AN ANALYSIS OF TEXAS RAINFALL DATA AND ASYMPTOTIC PROPERTIES OF SPACE-TIME COVARIANCE ESTIMATORS

A Dissertation

by

BO LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2006

Major Subject: Statistics

AN ANALYSIS OF TEXAS RAINFALL DATA AND ASYMPTOTIC PROPERTIES OF SPACE-TIME COVARIANCE ESTIMATORS

A Dissertation

by

BO LI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Michael Sherman
	Raymond J. Carroll
Committee Members,	Bani Mallick
	Marian Eriksson
Head of Department,	Simon J. Sheather

August 2006

Major Subject: Statistics

ABSTRACT

An Analysis of Texas Rainfall Data

and Asymptotic Properties of Space-time Covariance Estimators. (August 2006)

Bo Li, B.S., Shanghai Jiao Tong University, P. R. China;

M.S., Shanghai Jiao Tong University, P. R. China;

M.S., Texas A&M University

Co-Chairs of Advisory Committee: Dr. Michael Sherman
Dr. Raymond J. Carroll

This dissertation includes two parts. Part 1 develops a geostatistical method to calibrate Texas NexRad rainfall estimates using rain gauge measurements. Part 2 explores the asymptotic joint distribution of sample space-time covariance estimators. The following two paragraphs briefly summarize these two parts, respectively.

Rainfall is one of the most important hydrologic model inputs and is considered a random process in time and space. Rain gauges generally provide good quality data; however, they are usually too sparse to capture the spatial variability. Radar estimates provide a better spatial representation of rainfall patterns, but they are subject to substantial biases. Our calibration of radar estimates, using gauge data, takes season, rainfall type and rainfall amount into account, and is accomplished via a combination of threshold estimation, bias reduction, regression techniques and geostatistical procedures. We explore a varying-coefficient model to adapt to the temporal variability of rainfall. The methods are illustrated using Texas rainfall data in 2003, which includes WAR-88D radar-reflectivity data and the corresponding rain gauge measurements. Simulation experiments are carried out to evaluate the accuracy

of our methodology. The superiority of the proposed method lies in estimating total rainfall as well as point rainfall amount.

We study the asymptotic joint distribution of sample space-time covariance estimators of stationary random fields. We do this without any marginal or joint distributional assumptions other than mild moment and mixing conditions. We consider several situations depending on whether the observations are regularly or irregularly spaced, and whether one part or the whole domain of interest is fixed or increasing. A simulation experiment illustrates the asymptotic joint normality and the asymptotic covariance matrix of sample space-time covariance estimators as derived. An extension of this part develops a nonparametric test for full symmetry, separability, Taylor's hypothesis and isotropy of space-time covariances.

ACKNOWLEDGEMENTS

I feel so lucky to have Dr. Sherman and Dr. Carroll as my advisors in the most important stage in my life. They guided me during my research from the very early stages to writing the technical report. They showed me what enthusiastic and honest researchers are. They encourage me when I feel frustrated, and they truly care about my well-being. They are more than my advisors, they are the people that I can always trust and ask for advice for whatever reason. I wish I could have found a better word than "Thank you" to convey my appreciation for them.

I thank Dr. Bani Mallick and Dr. Marian Eriksson for their willingness to serve on my committee. Many thanks to Dr. Genton for offering me help in many situations. I also want to thank Dr. Cline for being my master's advisor and thank Dr. Longnecker for encouraging me to pursue my Ph. D. degree. I thank Dr. Carozza for currently supporting my study.

I thank my parents for raising and supporting me for so long. Many years ago, they taught me how to smile, how to say "thank you", and they taught me right from wrong. They devoted their life to their children. Now, they are old and I grew up. I especially thank my mom for her courage and the strength she shows while fighting with cancer. She always endows me with strength and encourages me to overcome difficulties. To me, she is the most wonderful mom in the world. I also thank my sister and brother for having taken care of me since I was a child.

I thank my husband, Yu, for his support and encouragement. His brightness, optimism and humor make life full of joy. Thanks to all my friends who have ever given me help.

TABLE OF CONTENTS

Page

ABSTRACT iii

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER

I INTRODUCTION 1

1.1 Random Fields and Stationarity 1
1.2 Geostatistics 2
1.3 Spatio-Temporal Processes and the Homogeneous Poisson Process 5
1.4 Overview Structure 7

II A GEOSTATISTICAL METHOD FOR TEXAS NEXRAD DATA CALIBRATION 8

2.1 Introduction 8
2.2 Geostatistical Method 11
2.3 Results 19
2.4 A Simulation 31
2.5 Discussion 36

III ON THE ASYMPTOTIC JOINT DISTRIBUTION OF SAMPLE SPACE-TIME COVARIANCE ESTIMATORS 39

3.1 Introduction 39
3.2 Discrete Lags 42

CHAPTER Page

3.3 Continuous Lags 45
3.4 Simulation 49

IV CONCLUSION AND FUTURE RESEARCH 53

4.1 Introduction 53
4.2 Various Assessments for Covariances 55

REFERENCES 59

APPENDIX A 64

APPENDIX B 66

VITA 83

LIST OF TABLES

TABLE Page

1 Simulation results for a 3×3 grid, $\phi = 1$ 51

2 Simulation results for a 3×3 grid, $\phi = 1.5$ 52

3 Simulation results for a 4×4 grid, $\phi = 1$ 52

LIST OF FIGURES

FIGURE		Page
1	Illustration of the moving window	17
2	The 60 weather stations	20
3	The NexRad grid	21
4	Details of NexRad showing 4 counties highlighted in Fig. 3	21
5	The 311 rain gauges	22
6	Daily threshold estimates	23
7	$\widehat{\beta}_0(t)$ with respect to $\widehat{\beta}_2$	25
8	$\widehat{\beta}_1(t)$ with respect to $\widehat{\beta}_2$	26
9	A plot of residuals vs. predicted value	27
10	An example of the semivariogram from the regression residuals	28
11	A comparison between predictors on 4/23/2003	30
12	A comparison between predictors on 7/6/2003	31
13	Simulated NexRad random field over 224×224 grid	35
14	Simulated gauge random field	35
15	Prediction of gauge random field	36
16	Partition of the random field for Lemma B.1.2	68
17	Partition of the random field for Lemma B.1.4	72

CHAPTER I

INTRODUCTION

Spatial statistics can answer not only the "how much" question, but also the "how much is where" question due to the fact that observations in close spatial proximity tend to be more similar than is expected for observations that are more spatially separated. This dissertation discusses two problems in spatial statistics. In Chapter II, we develop a geostatistical method to analyze Texas rainfall data, and in Chapter III, we explore the asymptotic joint distribution of sample space-time covariance estimators. The rest of this Chapter gives the background behind Chapter II and III. Section 1.1 introduces the random field and the concept of stationarity. Section 1.2 addresses basic geostatistical concepts. Section 1.3 turns to spatio-temporal processes and the Poisson process, and Section 1.4 describes the structure of this dissertation.

1.1 Random Fields and Stationarity

A random field, $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$, $d \geq 1$, also called a random process, is a family or collection of random variables, the members of which can be identified or located (indexed) according to some metric. To make inference from a random field possible, we have to make assumptions about the random field, otherwise, the data represents an incomplete sampling of a single realization (Cressie, 1993, p. 53). A weak assumption of a random field is based on the expected value of $Z(\mathbf{x})$ and the existence of var$\{Z(\mathbf{x})\}$. A random field under this assumption is called weakly (second-order)

The format and style follow that of *Biometrics*.

stationary if

$$\mathrm{E}\{Z(\mathbf{x})\} = \mu, \text{ for all } \mathbf{x} \in \mathbb{R}^d,$$

and

$$\mathrm{cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\} = C(\mathbf{x}_1 - \mathbf{x}_2), \text{ for all } \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d,$$

where μ is a constant and the function $C(\cdot)$ is called a covariance function. This latter assumption ensures that the covariance function is estimable. In Chapter II, we assume the random process of regression residuals is second-order stationary and thus we are able to make the optimal prediction on the unobserved residuals using geostatistical methods.

A stronger notion of stationarity is called strict (strong) stationarity or homogeneity if its probabilistic structure is invariant under spatial translations. Specifically, for any set $E \subset \mathbb{R}^d$, and for any spatial lag $\mathbf{k} \in \mathbb{R}^d$, the joint distribution of the random variables $\{Z(\mathbf{x}), \mathbf{x} \in E\}$ is identical to the joint distribution of $\{Z(\mathbf{x}), \mathbf{x} \in E + \mathbf{k}\}$, where $E + \mathbf{k} = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \mathbf{k}, \text{ with } \mathbf{x} \in E\}$. In Chapter III, we assume the random field is strictly stationary in order to make inferences on covariance estimators.

1.2 Geostatistics

Geostatistics is a collection of statistical methods which are applied to geostatistical data. Geostatistical data is characterized by the domain of the spatial data. Let $\{Z(\mathbf{x}), \mathbf{x} \in D \subset \mathbb{R}^d\}$ be a random process over a domain D. If $Z(\mathbf{x})$ can be observed everywhere within D and the locations of observations are non-stochastic, then the observations over D are geostatistical data. There are numerous geostatistical data in multiple disciplines, such as the wolfcamp-aquifer data and soil-water tension data in Cressie (1993). Our rainfall data in Chapter II is also geostatistical data, since

the locations of rainfall observations are not random. The objective of a statistical inquiry with such data is the spatial distribution of the attribute Z, and this usually involves exploring the spatial correlation using the sample data.

In geostatistical applications, it is common to work with the semivariogram rather than the covariance function. The semivariogram is defined by Matheron (1963) as

$$\gamma(\mathbf{x}_1 - \mathbf{x}_2) = \frac{1}{2}\mathrm{E}[\{Z(\mathbf{x}_1) - Z(\mathbf{x}_2)\}^2]. \tag{1.1}$$

The function $2\gamma(\cdot)$ is called the variogram. If $Z(\cdot)$ is second-order stationary, the semivariogram can be expressed in terms of the covariance function by

$$\gamma(\mathbf{x}_1 - \mathbf{x}_2) = C(\mathbf{0}) - C(\mathbf{x}_1 - \mathbf{x}_2).$$

Some other terminologies widely used in geostatistics and related to the variogram include nugget, sill, and range. The variogram may exhibit an apparent discontinuity at the origin. The magnitude of this discontinuity is called the nugget. With no nugget effect, $\gamma(\mathbf{0}) = 0$. The sill is the maximum value of the semivariogram if it exists, and the range is the distance at which the variogram attains the sill. If the semivariogram achieves the sill only asymptotically, such as the semivariogram of the Matérn class, then the practical range is typically defined as the lag distance at which the semivariogram achieves 95% of the sill.

Smoothness is another important property of the random process $Z(\cdot)$ that needs to be addressed, especially when we choose the Matérn class (Handcock and Wallis, 1994) as the covariance structure, as we do in Chapter II. Here we define the smoothness by mean square differentiability. A process Z on $\mathbb{R}$ with finite second moments is called mean square differentiable at t if $\{Z(t + h_n) - Z(t)\}/h_n$ converges in L^2 for all sequences $\{h_n\}$ converging to 0 as $n \to \infty$ with limit independent of $\{h_n\}$. If such a limit exists, we call it $Z'(t)$. To define higher-order mean square derivatives,

we say Z is m-times mean square differentiable if it is $(m-1)$-times mean square differentiable and $Z^{(m-1)}$ is mean square differentiable.

The smoothness of a random field plays a critical role in interpolation problems (Stein, 1999). Furthermore, there is often no basis for knowing a priori the degree of smoothness of some physical process modeled as a random field. Thus, it is prudent to use classes of models that allow for the degree of smoothness to be estimated from the data rather than restricting the smoothness a priori. The Matérn model allows for great flexibility in the smoothness of the random field while keeping the number of parameters manageable. The critical parameter in the Matérn class is the smoothness parameter ν. The larger ν is, the smoother the random field. In particular, $Z(\cdot)$ is m times mean square differentiable if and only if $\nu > m$. However, many other models assume a specific amount of smoothness as one of the attributes of the model. For example, an exponential model is a special case of the Matérn class with $\nu = 0.5$.

A popular method for interpolating geostatistical data is "kriging", which is synonymous with "optimal prediction" in the sense of minimizing mean-squared prediction error. Among the several versions of kriging, "ordinary kriging" is a common procedure, which requires the following two assumptions.

1. Model assumption:

$$Z(\mathbf{x}) = \mu + \epsilon(\mathbf{x}),\ \mathbf{x} \in D,\ \mu \in \mathbb{R} \text{ and } \mu \text{ unknown},$$

   where the correlated error process $\epsilon(\cdot)$ satisfies $\mathrm{E}\{\epsilon(\cdot)\} = \mathbf{0}$ and $\mathrm{var}\{\epsilon(\cdot)\} = \Sigma$ for some matrix Σ.

2. Predictor assumption:

   Let $\mathbf{Z}_n = (Z(\mathbf{x}_1), ..., Z(\mathbf{x}_n))^{\mathrm{T}}$ be observations at spatial locations $(\mathbf{x}_1, ..., \mathbf{x}_n)$,

$$\widehat{Z}(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i), \quad \sum_{i=1}^{n} \lambda_i = 1.$$

The goal is to choose weights $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_n)^{\mathrm{T}}$ that minimize

$$\mathrm{E}[\{\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{Z}_n - Z(\mathbf{x}_0)\}^2] \text{ subject to } \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{1} = 1,$$

where $\mathbf{x}_0$ is an arbitrary location. The results are

$$(\lambda_1, \cdots, \lambda_n, m)^{\mathrm{T}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_{\mathbf{0}},$$

$$\text{where } \boldsymbol{\gamma}_0 \equiv (\gamma(\mathbf{s}_0 - \mathbf{s}_1), \cdots, \gamma(\mathbf{s}_0 - \mathbf{s}_n), 1)^{\mathrm{T}}, \text{ and}$$

$$\boldsymbol{\Gamma}_{i,j} = \begin{cases} \gamma(\mathbf{s}_i - \mathbf{s}_j) & \text{if} \quad i, j = 1, \cdots, n, \\ 1 & \text{if} \quad i = n+1, j = 1, \cdots, n, \\ 1 & \text{if} \quad j = n+1, i = 1, \cdots, n, \\ 0 & \text{if} \quad i = n+1, j = n+1, \end{cases}$$

and m denotes a Lagrange multiplier that ensures $\sum_{i=1}^{n} \lambda_i = 1$. Under squared error loss, the ordinary kriging predictor is the best linear unbiased predictor. Moreover, it is the best unbiased predictor if $Z(\cdot)$ is a Gaussian process. We employ ordinary kriging to interpolate regression residuals in Chapter II.

1.3 Spatio-Temporal Processes and the Homogeneous Poisson Process

Many spatial processes change over time and thus addressing the time component in space-time processes must not be overlooked. When spatial data are collected over time, a spatio-temporal statistical analysis can provide benefits not possible from a spatio-only approach. Haas (1995) developed a prediction method based on seasonal rainfall-deposited sulfate over the conterminous United States between summer 1986 and summer 1992. de Luna and Genton (2005) presented a family of predictive spatio-temporal models for spatially sparse but temporally rich data.

Unfortunately, statistical tools for the analysis of spatio-temporal processes are not as fully developed as methods for time series or spatial data alone. The temptation

arises to separate spatial analysis and temporal analysis, and this allows predictions in space or time only, since interpolation of observations in a continuous space-time process should take into account the interactions between the spatial and temporal components. For this reason, joint analysis of spatio-temporal data are preferable to separate analyses. However, in the process of building a joint model, separate analyses are valuable tools. The random field is said to have separable spatio-temporal covariance C if C can be factored into the production of a purely spatial and a purely temporal covariance function C_S and C_T, respectively (e.g., Schabenberger and Gotway, 2005).

Exploring the separability of spatio-temporal covariance functions motivates us to study the asymptotic distribution of sample space-time covariance estimators as in Chapter III. Let $\{Z(\mathbf{s},t), \mathbf{s} \in S \subset \mathbb{R}^2, t \in T \subset \mathbb{R}\}$ be a space-time process over domain $S \times T$. We consider several situations including the case when observations are irregularly spaced in S or irregularly spaced in $S \times T$. In those particular cases, we assume the points at which Z is observed in the irregularly spaced domain are random in number and location (e.g. Karr, 1986). Specifically they are the points of a homogeneous multi-dimensional Poisson process N.

What is a homogeneous Poisson process? Let's start with the point process to answer this question. Informally speaking, a point process is a stochastic model governing the location of events in some set. The quantity $\nu(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^{\mathbf{d}}$ of a point process that measures the average number of events per unit area centered at location $\mathbf{x}$ is called the intensity. $\nu(\mathbf{x})$ is defined as a limit (e.g. Schabenberger and Gotway, 2005) since it is considered a function of points on an area basis,

$$\nu(\mathbf{x}) = \lim_{\lambda(d\mathbf{x})\to 0} \frac{\mathrm{E}\{N(d\mathbf{x})\}}{\lambda(d\mathbf{x})}$$

where $d\mathbf{x}$ denotes an infinitesimal disc in $\mathbb{R}^d$ centered at $\mathbf{x}$, $N(\cdot)$ is the total number

of events within $\cdot$ and $\lambda(\cdot)$ is the Lebesgue measure of $\cdot$. A point process is referred to as the homogenous Poisson process if the following criteria are met: The intensity ν is homogeneous throughout the domain of interest ($\nu(\mathbf{x}) = \nu$, say), the number of events in two non-overlapping subregions A_1 and A_2 are independent, and the number of events in any subregion has a Poisson distribution. Thus in a homogeneous Poisson process, events are distributed uniformly and independently throughout the domain.

1.4 Overview Structure

Chapter II develops a geostatistical method to calibrate the Texas NexRad rainfall estimates using rain gauge measurements. This chapter mainly contains a methodological proposal, a data analysis employing this method, and a simulation experiment evaluating the methodology. Chapter III is devoted to the study of the asymptotic joint distribution of sample space-time covariance estimators. This chapter focuses on derivations and presents a small simulation to verify the theoretical results. Chapter IV considers a future research topic of developing a nonparametric test for full symmetry, Taylor's hypothesis, isotropy and separability of spatio-temporal covariance functions based on the results in Chapter III. Lemmas and proofs of the theorems are detailed in the appendices.

CHAPTER II

A GEOSTATISTICAL METHOD FOR TEXAS NEXRAD DATA CALIBRATION

2.1 Introduction

Precipitation is one of the most important hydrologic model inputs and is characterized by spatial and temporal variability. Traditionally, point precipitation measurements at rain gauges have been used with hydrological models. Since rain gauges physically measure the depth of rainfall at the points of measurement, they generally provide good quality data. However, rain gauge networks are usually too sparse to capture the spatial variability of precipitation over the hydrologic system. This problem becomes more critical when simulating large river basins.

Compared to rain gauges, weather radars such as the Next generation weather Radar (NexRad), formally known as the Weather Surveillance Radar-1988 Doppler (WSR-88D) of the United States provide precipitation data with much better spatial and temporal sampling frequencies. A better representation of rainfall variability can be accomplished in a hydrologic model by using radar rainfall data. However, there are several possible sources of errors and possibilities of "data contamination" in the radar estimates of precipitation, as the radar estimates are not the real measurement of rainfall and thus subject to errors (Jayakrishnan, Srinivasan, and Arnold, 2004). For this reason, we seek to improve the quality of radar precipitation data over the study area using rain gauge measurements.

Radar data are formed over three stages. Originally radar measures the reflectivities by volume scans over a fixed polar grid with a radial resolution of one degree in azimuth by 1 km in range. These reflectivities are then converted into rainfall rates by the convective Z-R relationship and the Rosenfeld tropical Z-R relationship,

and further converted to rainfall depth on a grid, called the Stage I output. Stage II processing corrects Stage I output for the individual radar using a bias adjustment factor which is calculated as the ratio of the sum of all positive rain gauge data over a specific radar umbrella to the sum of all non-zero Stage I gridded output at the same gauge locations over the same spatio-temporal window of sampling (`http://www.srh.noaa.gov/wgrfc/resources/projects/stageiii_paper/default.html`). Finally, Stage III processing mosaicks the data from multi-radar for the areas under the umbrella of more than one radar .

In this article, we propose a methodology to calibrate daily rainfall data. Our goal is to improve the quality of radar precipitation data using rain gauge data based on Texas rainfall data set in 2003. There are three sets of daily data: one is station data which is of high quality, but contains only 60 weather stations; another one is gauge data that is of relatively good quality and contains 664 rain gauges; the last one is daily NexRad estimates over a 4×4 km grid over Texas. The gauge data is compiled at the National Climate Data Center (NCDC) of National Weather Service (NWS), and the NexRad data is Stage III WSR-88D precipitation data from NWS. Both station data and NexRad data give daily rainfalls from 7am to 7am, but not all of the gauges are collected from 7am to 7am. Only 311 out of 664 are collected at 7am and are thus comparable with the station and NexRad data.

Barancourt, Creutin, and Rivoirard (1992) proposed a geostatistical scheme for modelling rainfall using two mutually independent random fields, $Z(\mathbf{s}) = I(\mathbf{s})F(\mathbf{s})$, where $\mathbf{s}$ denotes location. The first is a binary random field which models the intermittency of the rainfall. Indicator kriging gives a map of rain probabilities which can be converted by thresholding into a contour of the rainy areas. Inside the rainy areas, the second random field accounts for the variability of the nonzero rainfall value, and it is also estimated by kriging. Their method is useful when only one data

set is available and the interest focuses on delineation of the rainy area. If the mean areal rainfall over a fixed and large domain is desired, Barancourt's method does no better than spatial kriging. De Oliveira (2004) introduced an analogous model for short time rainfall based on the separate modelling of spatial occurrence of rainfall and the spatial distribution of positive rainfalls. This model is essentially the same as Banrancourt's but is formulated differently.

Gel, Raftery, and Gneiting (2004) described a geostatistical output perturbation method to deal with two sets of temperature data. This situation is similar to ours in that we both have less accurate values on a regularly spaced grid and more accurate values at irregularly spaced locations, and share the same goal of finding a methodology to improve the performance of less accurate values. However, our situation differs from theirs in several ways, principally the rainfall data is a mixture of hard zeroes and continuous nonzero values while the temperature data are absolutely continuous.

We develop a geostatistical method combined with threshold estimation, bias reduction and regression techniques to calibrate the NexRad estimates using the rain gauge measurements. To account for hard zeroes, we include a binary random variable in the model to delineate the dry and wet areas based on the NexRad estimates. We apply regression to reduce the bias of NexRad estimates and obtain the spatially correlated errors. We model and estimate the errors using a geostatistical method, then finally make predictions with the bias reduced NexRad estimates and the estimated errors. In order to adapt to the variability of rainfall over the time, we employ the varying-coefficient model proposed by Hastie and Tibshirani (1993). Our proposed method preserves the spatial occurrence pattern and spatial variability within the wet areas of the true rainfall.

In Section 2.2, we describe the geostatistical method, including the basic statistical model, parameter estimation, prediction using the estimated model, and predic-

tion validation to evaluate our method and assess our predictions. In Section 2.3, we apply our method to the NexRad estimates and the corresponding gauge measurements on Texas in 2003, and give the prediction validation results. In Section 2.4, we describe the simulation experiments and compare the simulation results with those of the data analysis. Finally, in Section 2.5, we discuss the features of the proposed method and possible future work.

2.2 Geostatistical Method

We now describe our geostatistical method. First we outline our underlying statistical model. Then we describe how to estimate model parameters from data and how to use the estimated model to make predictions. Finally, we propose a validation method to assess the accuracy of our model.

2.2.1 Statistical model

Although the 60 stations are the most reliable, they are too sparse to capture the spatial features of the rainfall data, but they will be helpful to assess the quality of our predictions. The 311 gauges are dense enough to give us some spatial information, so we use the 311 gauges to calibrate the NexRad data.

Let X_i denote the gauge measurement, $i = 1, ..., 311$, W_j denote NexRad estimates, $j = 1, ..., 50151$. Let s_i denote the location of X_i, s'_j denote the location of W_j. For s'_j close to s_i, we expect that the value of W_j is close to the value of X_i if they are recorded on the same day. For W_j with $\min_j \|s'_j - s_i\|$, define $W_j = W_i$. There are four cases in the data:

(1) $X_i = 0$, $W_i = 0$;

(2) $X_i > 0$, $W_i = 0$;

(3) $X_i = 0$, $W_i > 0$;

(4) $X_i > 0$, $W_i > 0$.

Cases (1) and (4) are what we hope to see in the data, while cases (2) and (3) are not. The goal is to calibrate W using X. If we use $W_i = 0$ as the classifier for $X_i = 0$, we misclassify the X_i's in cases (2) and (3). To reduce the misclassification error, we introduce a new classifier β_2 such that the four cases become:

(1) $X_i = 0$,$W_i \leq \beta_2$;

(2) $X_i > 0$,$W_i \leq \beta_2$;

(3) $X_i = 0$,$W_i > \beta_2$;

(4) $X_i > 0$,$W_i > \beta_2$.

We expect the introduction of β_2 to reduce the misclassification rate. For NexRad data that are above the threshold, we apply linear regression to remove bias from the NexRad data. In order to adapt to the possible temporal variability of rainfall, we employ a varying-coefficient model to allow the coefficients to vary with time.

Based on the above discussion, we construct the following model:

$$X^{1/k} = \{\beta_0(t) + \beta_1(t)W^{1/k} + \xi(W, \boldsymbol{\beta})\epsilon\} I\{W > \beta_2(t)\} \tag{2.1}$$

where X and W are random processes of rain gauge and NexRad. The index t denotes time, and serves as an effect modifying variable (Hastie and Tibshirani, 1993), while $I(\cdot)$ denotes an indicator function and $\beta_2(t)$ is a time-dependent threshold. The functions $\beta_0(t)$ and $\beta_1(t)$ are time-dependent additive and multiplicative bias terms. Let $\boldsymbol{\beta}$ denote $(\beta_0(t), \beta_1(t))^{\mathrm{T}}$ and $\xi(W, \beta)$ is a function to account for nonconstant variance. The spatial errors $\epsilon \sim F(\mathbf{0}, \Sigma)$, where F is a multivariate distribution function with covariance matrix Σ. We parameterize Σ from the Matérn class of covariance functions. We follow the parameterization recommended by Handcock and Wallis (1994) for the Matérn covariance function:

$$K(r) = \frac{\sigma}{2^{\nu-1}\Gamma(\nu)}(\frac{2\nu^{\frac{1}{2}}r}{\rho})^{\nu}\mathcal{K}_{\nu}(\frac{2\nu^{\frac{1}{2}}r}{\rho}), \tag{2.2}$$

where $\Gamma(\cdot)$ is the gamma function; $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν; σ is the variance, also called the sill in geostatistics; ν is the smoothness parameter. The larger ν is, the smoother the random process. The parameter ρ measures how quickly the correlations of the random field decay with distance. In this parameterization, it has the attractive feature that its interpretation is largely independent of ν.

The power $1/k$ is used to make data closer to Gaussian, since we seek to use likelihood methods to fit model parameters. Although the log transformation is one common way to transform rainfall data to normal, and it demonstrates good performance for the accumulated rainfall amount, it is inappropriate for point estimates due to the hard zeroes in the data set. Adding a small constant to the zero value allows the log transform, but different constants lead to different parameter estimates. Our power transform avoids this choice.

This model is the product of a Gaussian random field with linear mean function and a binary random field which is independent of the former random field. It is related to the model of Barancourt et al. (1992) and De Oliveira (2004). They proposed a model as a product of two independent random fields $Z(s) = R(s) \times I(s)$. This kind of model has the flexibility to reproduce the spatial association structure of rainfall fields, because the spatial association structure of the rainfall occurrence and rainfall amounts process are governed by the different association structures. One drawback of their method being applied to our data is the lack of smooth transitions between the dry and wet sub-regions which are the norm for most rainfall patterns. Moreover, their procedures are mainly developed for modeling the rainfall random field using one data set, so it is often not efficient for calibrating one data set using another data set. However our threshold, which is estimated directly from the NexRad data, avoids this problem by following the true transition patterns from dry to wet area

whether the transition is smooth or not.

2.2.2 Parameter estimation

We estimate model parameters given by (2.1) and (2.2) using data from 2003. The sequence of parameters being estimated is $k, \beta_2(t), \beta_0(t), \beta_1(t)$ followed by the parameters involved in Σ of $\text{var}(\epsilon)$. If weighted least squares are needed, $\xi(W, \beta)$ will be estimated simultaneously with $\beta_0(t)$ and $\beta_1(t)$.

We choose an appropriate k based on the Shapiro-Wilk test of the transformed gauge data. We estimate the time-dependent coefficients by the plane fitting procedure described by Hastie and Tibshirani (1993), noting that our varying-coefficient model contains only a single effect modifying variable t. The main idea of the plane fitting procedure is to construct a neighborhood for a specific t_0 in the (t, W)-plane, fit the model of X on W using the data in the neighborhood of t_0 to obtain point estimates of $\widehat{\beta}_2(t_0)$, $\widehat{\beta}_0(t_0)$ and $\widehat{\beta}_1(t_0)$, and repeat this for different t_0.

To choose $\beta_2(t)$, we pair the gauge and the average of the nearest 4 NexRad neighbors, we then employ logistic regression to estimate the threshold by finding the decision boundary between $X = 0$ and $X > 0$ in terms of W when $W > 0$ (e.g., Hastie, Tibshirani, and Friedman, 2001). The reason for averaging is to make the NexRad data stable. Let $P(W) = \Pr(X > 0 \mid W > 0)$. The logistic regression model has the form

$$\text{logit}\{P(W)\} = \log\{\frac{P(W)}{1 - P(W)}\} = \alpha_0(t) + \alpha_1(t)W.$$

The threshold is the value of W for which the log-odds are zero, and this is the point defined by $\{W \mid \alpha_0(t) + \alpha_1(t)W = 0\}$, which leads directly to $\widehat{\beta}_2(t) = -\widehat{\alpha}_0(t)/\widehat{\alpha}_1(t)$. We explore several methods of estimating $\beta_0(t)$ and $\beta_1(t)$ and settle on linear regression on the transformed value of W_j.

Before we determine the form of $\xi(W, \boldsymbol{\beta})$, we first assess the heteroscedasticity of the residuals. Carroll and Ruppert (1988) suggested several methods of assessing heteroscedasticity based on a fairly large number of observations and varying degrees of data density. One is to compute the Spearman rank correlation coefficients of the absolute studentized residuals with the fitted values. Another way to cope with the density effect in large data sets while at the same time obtaining information about a model for the variability is to use nonparametric regression techniques. We carry out both in our analysis. If the analysis results detect heteroscedasticity, we employ weighted least squares to estimate β_0, β_1 and $\xi(W, \boldsymbol{\beta})$ iteratively.

Finally, we calculate residuals from the linear regression and find the underlying spatial pattern of the residuals. To estimate the variogram, we use daily residuals since the rainfall patterns differ from day to day. However calculating residuals at only the closest NexRad location whose rainfall value is above the threshold for each gauge, gives at most 311 residuals at a very limited number of spatial lags. This gives information which is too sparse to capture the spatial pattern. Instead we assume that the nearest 8 NexRad locations and the corresponding gauge have the same true rainfall (Jayakrishnan et al., 2004), and calculate the residuals at the nearest 8 NexRad locations for each gauge. This allows for a sufficient number of residuals at a variety of spatial lags to estimate the variogram.

We calculate:

$$\widehat{\epsilon}_{ij} = \frac{1}{\widehat{\xi}(W_{ij}, \widehat{\boldsymbol{\beta}})}\{X_i^{1/\widehat{k}} - \widehat{\beta}_0(t) - \widehat{\beta}_1(t) W_{ij}^{1/\widehat{k}}\},\ i = 1, ..., 311;\ 0 \leq j \leq 8,\ j \text{ is integer.} \tag{2.3}$$

where W_{ij} is the j^{th} nearest NexRad whose rainfall is above the threshold for the i^{th} gauge and $\widehat{\boldsymbol{\beta}} = \{\widehat{\beta}_0(t), \widehat{\beta}_1(t)\}^{\mathrm{T}}$.

It is known, e.g., Stein (1999), that using empirical semivariograms for model

selection can work disastrously for smooth processes. For this reason, we desire to use likelihood based methods. The full maximum likelihood method, however, tends to be prohibitively time consuming in this setting. We seek a simpler and faster procedure that approximates the full maximum likelihood method and works well in our implementation. Due to the well behaved transformed data and the fact that the spatial dependence of residuals decays quickly as distance between residuals increases, it's natural to use an approximate likelihood over moving windows (Stein, 1986, 1999). We center the moving windows on each gauge that is used to calculate the residuals in (2.3). Figure 1 shows an example of a moving window. The circle window is centered on the "*" which denotes a gauge location. The eight small circles around the "*" are the nearest eight NexRad locations for that gauge. The black dots belong to the nearest eight NexRad locations for some other gauges. We write the likelihood function using the residuals on all the dots and small circles within each window, and then pool up all the likelihood functions of all windows.

Specifically, the parameters in the variogram model are $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi, \nu)^{\mathrm{T}}$ which represent Matérn(nugget, sill, range, smoothness). Let K denote the total number of gauges that are involved in (2.3). Then, K is also the total number of moving windows. Let $\mathbf{e}_k$ denote the residual vector in the k^{th} window, $k = 1, 2, ..., K$, $\Sigma_k(\boldsymbol{\theta})$ denote the covariance matrix of $\mathbf{e}_k$, and let $|\cdot|$ denote the determinant of a matrix.

Our joint log likelihood is as follows up to a constant that does not depend on $\boldsymbol{\theta}$:

$$l(\boldsymbol{\theta}) = -\sum_{k=1}^{K}\{\log \mid \Sigma_k(\boldsymbol{\theta}) \mid + \mathbf{e}_k^{\mathrm{T}}\Sigma_k(\boldsymbol{\theta})^{-1}\mathbf{e}_k\}.$$

We estimate the variogram parameters by maximizing $l(\boldsymbol{\theta})$.

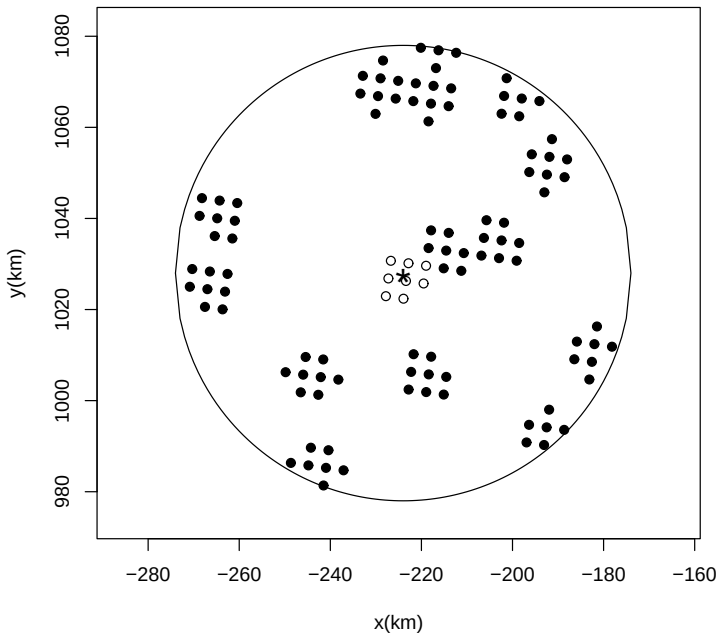

Figure 1: Illustration of the moving window

2.2.3 Prediction

Given model (2.1) and the estimated parameters, we now address the prediction at unknown locations on any given day.

First, we screen all the pairs of gauge and its nearest eight NexRad neighbors, ignore the ones that have $W_j < \widehat{\beta}_2(t)$ on that day, calculate the residuals $\widehat{\epsilon}$, the vector of residuals given in (2.3) using the appropriate parameters on all the pairs that have passed the screening.

Next, we employ our approximate maximum likelihood method over moving windows to estimate model parameters in the variogram on that day, and krige $\widehat{\epsilon}$ to location s_0 via ordinary kriging,

$$\widehat{\epsilon}(s_0) = \sum_{j=1}^{n} \lambda_j \widehat{\epsilon}_j, \tag{2.4}$$

where λ_j are the kriging weights.

Finally, we plug $\widehat{\epsilon}(s_0)$ into the following formula:

$$\widehat{X}(s_0) = ([\widehat{\beta}_0(t) + \widehat{\beta}_1(t)W(s_0)^{1/\widehat{k}} + \widehat{\xi}\{W(s_0), \widehat{\boldsymbol{\beta}}\}\widehat{\epsilon}(s_0)]^{\widehat{k}} - bias)I\{W(s_0) > \widehat{\beta}_2(t)\} \quad (2.5)$$

where *bias* is the estimated bias induced by the power transformation.

2.2.4 Validation

Among the 60 stations, 14 of them are collocated with gauges. The measurements at the remaining 46 station locations are not being used in our calibration, so we use these 46 stations to assess the quality of our prediction. We predict at each of the 46 stations, and compare the prediction with the station rainfall measurements by computing the statistics that are widely used by hydrologists, e.g., Jayakrishnan et al., 2004. Let Z_{il}, $i = 1, ..., 46$, $l = 1, ..., L$, denote the i^{th} station measurement on the l^{th} day, where L is the number of days being used in the comparison. Let $\widehat{Z}_{il}$ denote the predicted value of Z_{il}, $\widehat{Z}_{il} = \widehat{X}(s_{il})$ in (2.5), where s_{il} is the location of Z_{il}. Define $Z_l = \sum_{i=1}^{46} Z_{il}$, and $\widehat{Z}_l = \sum_{i=1}^{46} \widehat{Z}_{il}$. Z_l is the daily total of the station measurements and $\widehat{Z}_l$ is the predicted value of Z_l. Let $\bar{Z}$ denote the mean of the Z_l's. To assess accuracy, we calculate the:

1. Total Difference in precipitation, $D = \sum_{l=1}^{L}(\widehat{Z}_l - Z_l)$.
2. Estimation Bias (%), EB $= 100D/\sum_{l=1}^{L} Z_l$.
3. Estimation Efficiency, EE $= 1.0 - \dfrac{\sum_{l=1}^{L}(\widehat{Z}_l - Z_l)^2}{\sum_{l=1}^{L}(Z_l - \bar{Z})^2}$.
4. Sum of Square Prediction Error, SSPE $= \sum_{l=1}^{L}\sum_{i=1}^{46}(\widehat{Z}_{il} - Z_{il})^2$.

It is common to use NexRad as rainfall estimates in hydrology, for example, Bedient, Hoblit, Gladwell, and Vieux (2000) and Bedient, Holder, Benavides, and

Vieux (2003) utilized NexRad data in hydrological model for flood prediction. For this reason, we also compute the above four statistics using the nearest NexRad value as the estimate of the rainfall. Let W_{ij} denote the nearest NexRad for Z_{ij} on the j^{th} day. We formulate the corresponding $\widetilde{D}$, $\widetilde{\text{EB}}$, $\widetilde{\text{EE}}$ and $\widetilde{\text{SSPE}}$ for the nearest NexRad predictor by replacing $\widehat{Z}$ by W in 1-4. We compare our prediction with the nearest NexRad to see what (if any) improvement our method accomplishes.

2.3 Results

2.3.1 Data

To estimate the model and assess our predictions, we use rainfall data in Texas in 2003. There are three sources of daily rainfall data: one is of the highest quality, but contains only 60 weather stations, as shown in Figure 2; another is of good quality and contains 664 rain gauges; the last one is 50151 daily NexRad estimates over a 4×4 km grid as shown in Figures 3 and 4. Both high quality station data and NexRad data give daily rainfall from 7am to 7am, but not all of the good quality gauges are collected from 7am to 7am. Only 311 out of 664 are collected at 7am and are thus comparable with the high quality station and NexRad data. The spatial locations of the 311 gauges are shown in Figure 5. We use millimeter (mm) as the unit for rainfall and kilometer (km) as the unit for distance throughout.

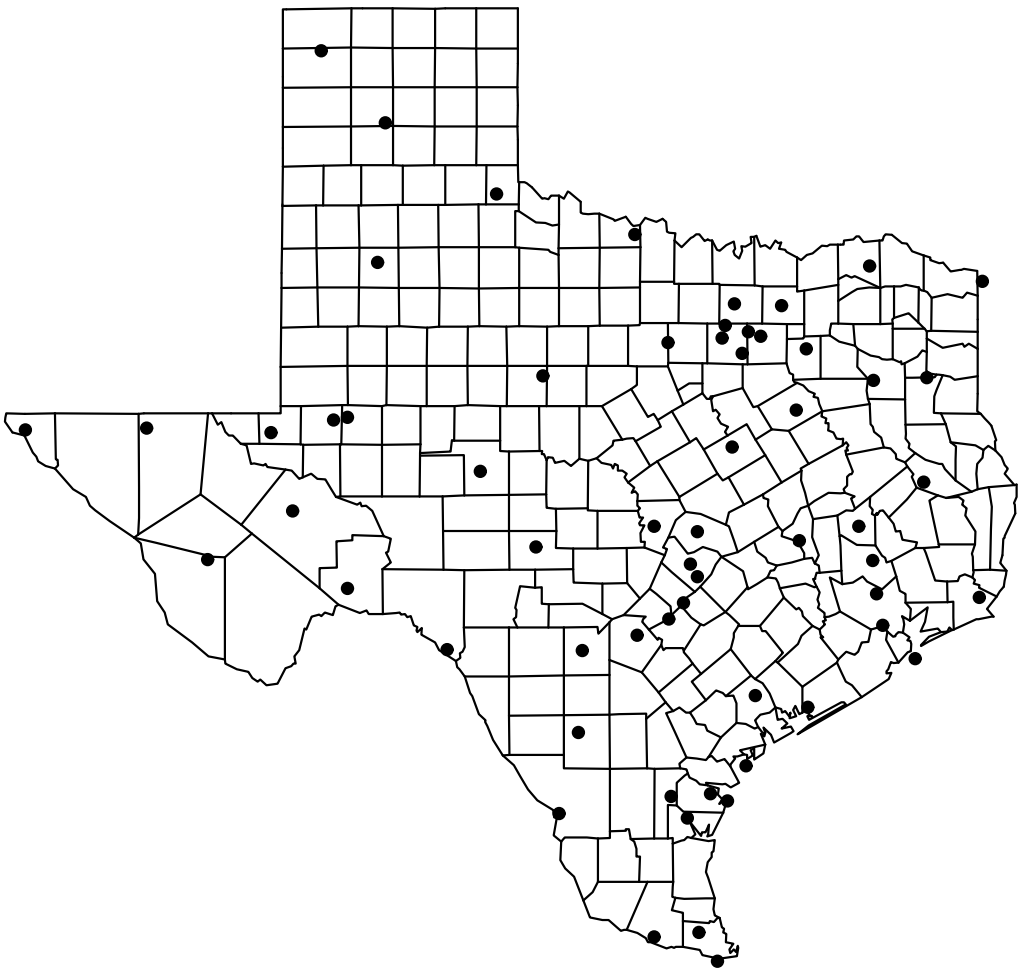

Figure 2: The 60 weather stations

2.3.2 Parameter estimation

We compare several power transformations with different powers. Using the Shapiro-Wilk test statistics of gauge data for different powers, we find that 1/3 is the best

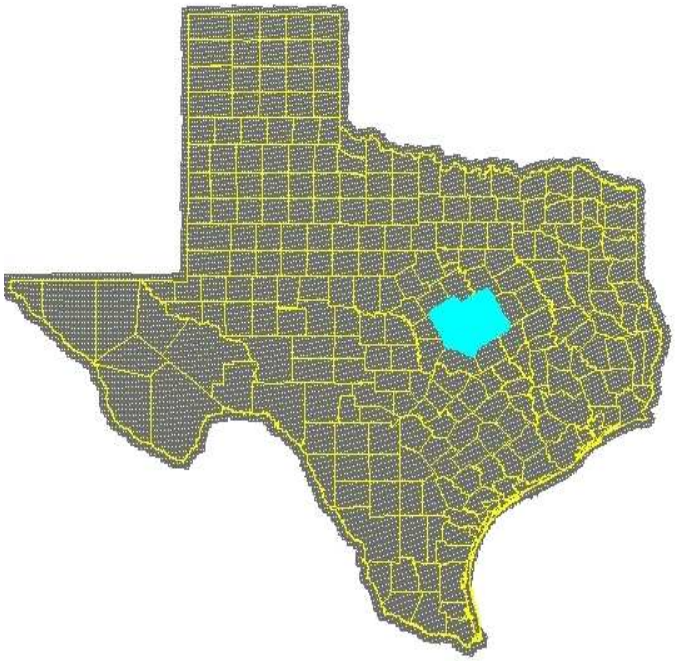

Figure 3: The NexRad grid

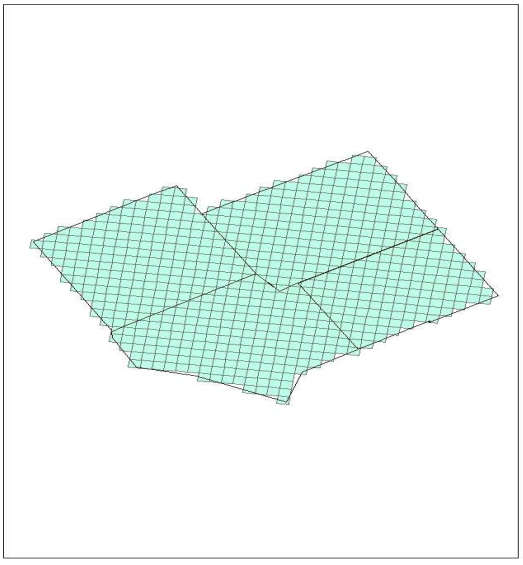

Figure 4: Details of NexRad showing 4 counties highlighted in Fig. 3

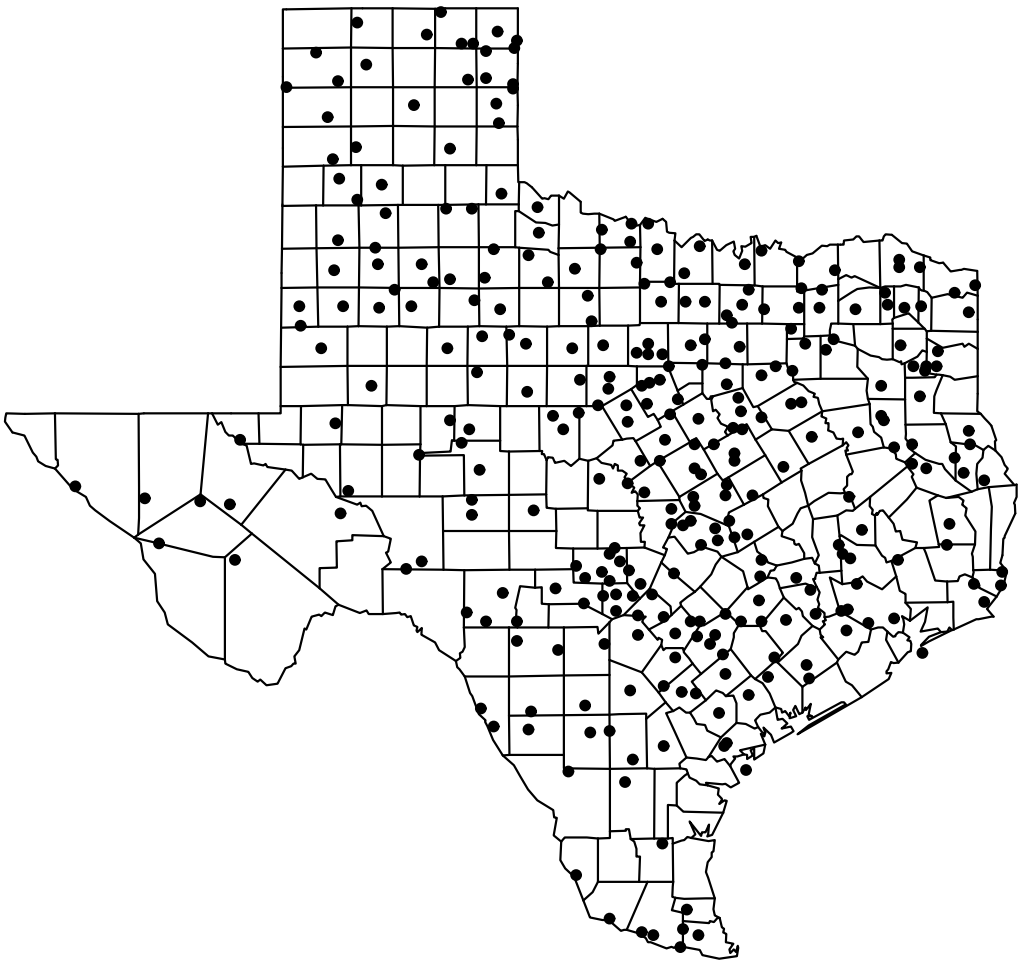

Figure 5: The 311 rain gauges

power among $1/k,\ k = 1, 2, ...$ for both seasons described below.

To estimate time-dependent coefficients by the plane fitting method discussed in Section 2.2.2, we choose t_0 as 7pm for each day and the neighborhood of t_0 as 24 hours

spanned from 7am to 7am, that is, we fit the model using daily data to obtain point coefficients estimates at discrete $t = 1, 2, ..., 365$. For each day, equivalently, for each discrete value of t, we select pairs (X, W) with $W > 0$ and fit a logistic regression model to the binary outcome of X in terms of rain or no rain. We calculate $\widehat{\beta}_2(t)$ for each t such that there are at least 20 pairs of $(X, W > 0)$ on the t^{th} day. Finally we obtain 219 point estimates of $\beta_2(t)$ together with their fitted cubic smoothing spline using 2 inner knots as depicted in Figure 6. With the observation that $\widehat{\beta}_2(t)$ is noisy, this picture shows only the estimates that are between -10 and 20 to make the main body of $\widehat{\beta}_2(t)$ clear. Using the unstable $\widehat{\beta}_2(t)$ as the corresponding threshold, we fit linear regression for each of 219 t's. We then find $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$ are even noisier, which suggests to combine daily data for threshold estimation.

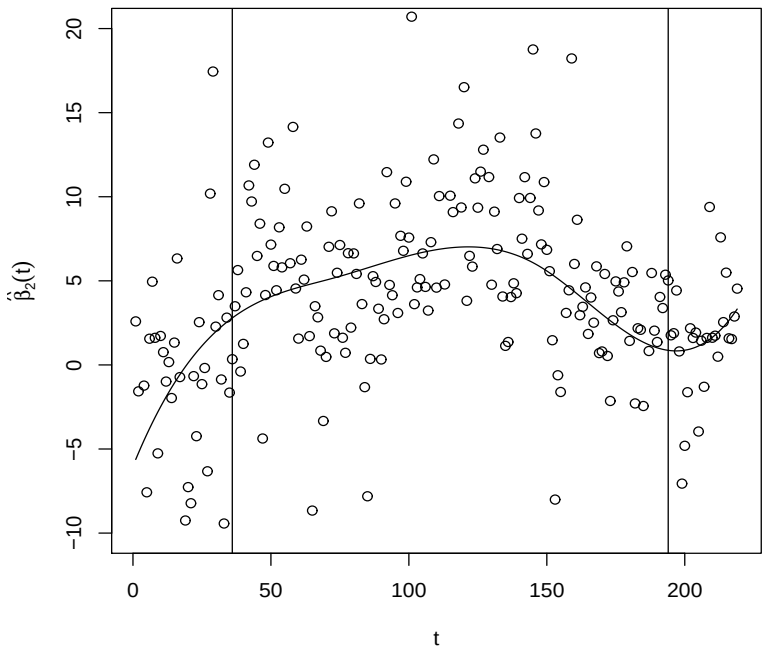

Figure 6: Daily threshold estimates

It has been pointed out, e.g., (Anagnostou, Krajewski, Seo, and Johnson, 1998),

that the rainfall is of qualitatively different types in the warm season (April-October) and the cold season (November-March). The warm season is dominated by convective rainfall, while the cold season is dominated by widespread stratiform rainfall. Convective rainfall is localized rainfall, like a thunderstorm. It is formed when cold currents from the north meet warm currents from the south. On the contrary, stratiform rainfall is more spread out and not focused on a few small areas. The two vertical lines shown in Figure 6 divide the whole plot into the cold season and the warm season. The cold season lie in the left area to the left line and the right area to the right line. The warm season lies in the area between the two vertical lines. It can be seen from Figure 6 that the estimates of $\beta_2(t)$ are approximately constantly high in the warm season and approximately constantly low in the cold season. This suggests estimating β_2 by combining the seasonal data. We get $\widehat{\beta}_2 = 0.5787$ in cold season and $\widehat{\beta}_2 = 5.6748$ in warm season. These two values roughly match the value of estimates displayed in Figure 6.

Let $|\cdot|$ denote the cardinality of $\cdot$, and M denote the total misclassification error rate.

$$M = \frac{|\, W \le \widehat{\beta}_2, X > 0 \,| + |\, W > \widehat{\beta}_2, X = 0 \,|}{|\, W, X > 0 \,| + |\, W, X = 0 \,|}.$$

The estimated thresholds reduce the misclassification error rates from 11% to 9.7% in the cold season, and from 15.9% to 11.3% in the warm season compared to using $W = 0$ as the threshold.

Given the specified $\widehat{\beta}_2$, we employ the plane fitting method again to estimate $\beta_0(t)$ and $\beta_1(t)$. In order to obtain a successful linear regression, we select 194 days and fit a linear model on the pairs $(X^{1/3}, W^{1/3} \mid W \ge \widehat{\beta}_2)$ on each of those days. The resulting $\widehat{\beta}_0(t)$ and $\widehat{\beta}_1(t)$ together with their fitted cubic smoothing splines using 2 inner knots are illustrated in Figures 7 and 8. The two vertical lines in Figure 7

play the same role as in Figure 6. $\widehat{\beta}_0(t)$ is then also approximately constant over each season, while $\widehat{\beta}_1(t)$ is approximately constant over the year. Nevertheless we estimate both β_0 and β_1 separately in the two seasons, as suggested by Anagnostou et al. (1998).

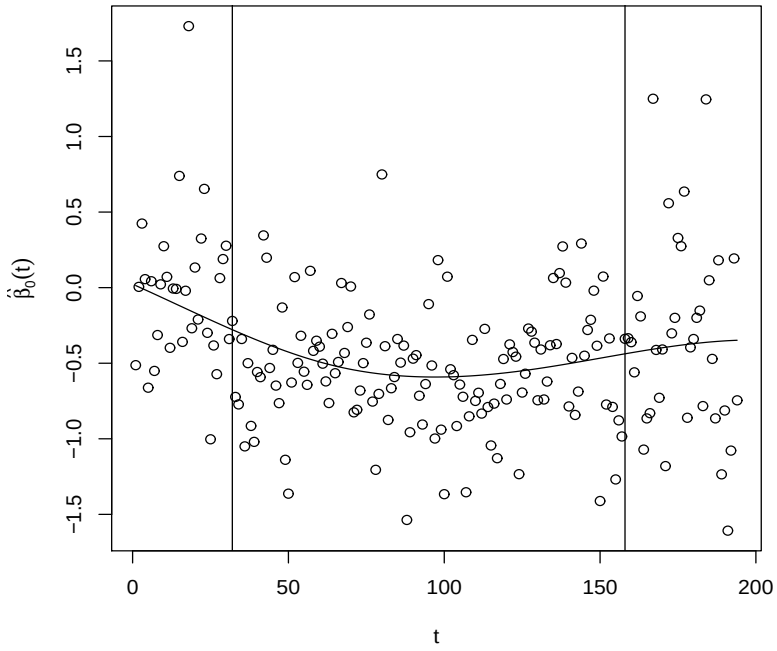

Figure 7: $\widehat{\beta}_0(t)$ with respect to $\widehat{\beta}_2$

We pool up the pairs $(X^{1/3}, W^{1/3} \mid W \geq \widehat{\beta}_2)$ within each season and fit them by a linear model, having assumed there is no rainfall for pairs with $W < \widehat{\beta}_2$, The data behave well after the cubic root transform, which allows us to use the Least Square (LS) regression. Using the seasonal LS regression, we find $R^2 = 0.3692$ in cold season and $R^2 = 0.2414$ in warm season. Being concerned about a possible quadratic relationship between NexRad and gauge, we fit a linear regression with an extra quadratic term in the model. We then obtain $R^2 = 0.3692$ in the cold season and $R^2 = 0.2415$ in the warm season. Thus there is little if any benefit to introduce a

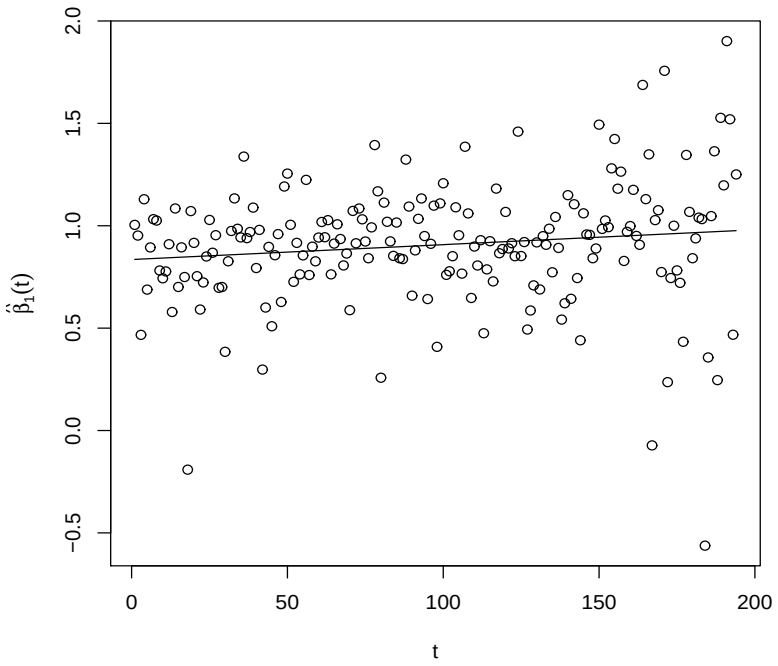

Figure 8: $\widehat{\beta}_1(t)$ with respect to $\widehat{\beta}_2$

quadratic term. The following are the LS parameters estimates using the gauge data and the averaged nearest 4 NexRad values.

Cold Season: $\widehat{\beta}_0^{LS} = -0.3143$, $\widehat{\beta}_1^{LS} = 0.938$;

Warm Season: $\widehat{\beta}_0^{LS} = -0.7174$, $\widehat{\beta}_1^{LS} = 1.0289$.

In the warm season $\widehat{\beta}_1$ is not significantly different from 1. This means there is little multiplicative bias in the warm season. These LS parameters estimates also approximately correspond with the values shown in Figures 7 and 8.

To assess possible heteroscedasticity, we use kernel regression with a global bandwidth to smooth the absolute studentized residuals. The bandwidth is selected by trial and error. The kernel regression shows the main body of standard deviation increases mildly from 0.8 to 1.0 as shown in Figure 9, so we don't weight residuals, that is, we set $\xi(W_i, \widehat{\beta}) = 1$. This is confirmed by the Spearman's rank correlation

test which gives $\widehat{\rho} = 0.02009846$, and a two sided p-value = 0.1374.

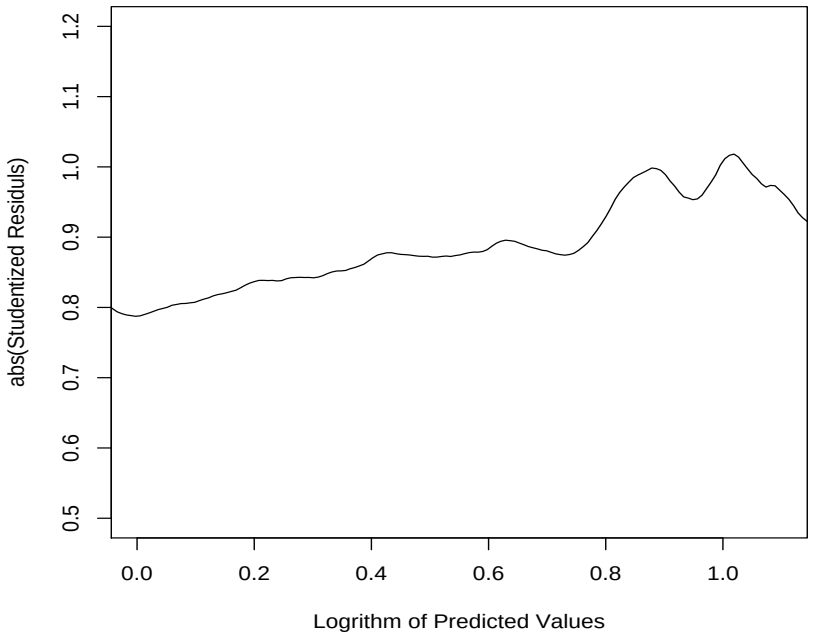

Figure 9: A plot of residuals vs. predicted value

Figure 10 shows the empirical semivariagrams of residuals on February 20th together with the fitted variogram model using an approximate likelihood method as illustrated in Section 2.2.2. Though the fitted model does not match the empirical variogram very well, we still decide to use this likelihood based estimate. Stein (1999) highly recommends using maximum likelihood in plug-in predictions even though it shows a worse match to the empirical variogram than moment based estimators. Viewing several semivariogram plots indicates it is not necessary to include a nugget term in the covariance model. In estimating variogram model parameters by the approximate likelihood method over moving windows, the radius of the moving window is set to 50km by examining the difference between the maximum likelihood estimated variogram models of different window radii and the required number of observations

within each window suggested by Stein (1999, p.172). In 2003, we selected 146 days that have the largest wet coverage to analyze the daily data. Two of the days do not have station data, and one day has too few NexRad locations whose values are above the threshold, so we remove those three days, and get the mean(std. dev.) of the sill, range and smoothness over the remaining 143 days: 1.023067(0.28544), 34.13421(13.90514), 0.858182(0.191535). The median of those parameters are (1.0, 31.07089, 0.8431755).

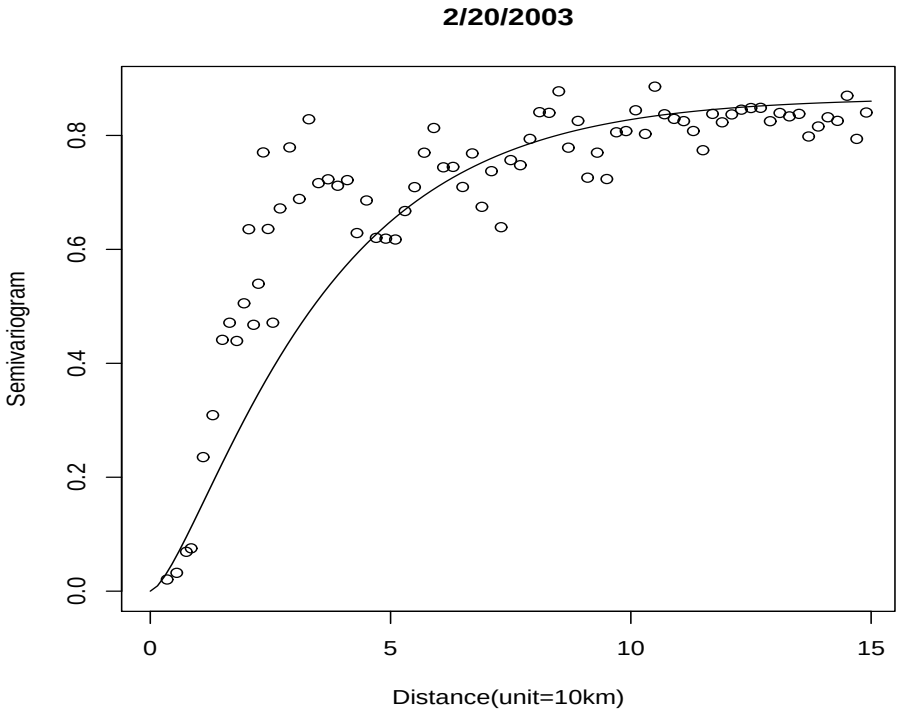

Figure 10: An example of the semivariogram from the regression residuals

2.3.3 Prediction and validation over rainy days

If we assume $\widehat{\beta}_0$ and $\widehat{\beta}_1$ in (2.5) are the true parameters, when $\xi\{W(s_0, \widehat{\beta})\} = 1$ and $\widehat{k} = 1/3$ (as obtained in Section 2.3.2), the bias in (2.5) can be estimated unbiasedly

by:

$$\widehat{\text{bias}} = 3\{\widehat{\beta}_0 + \widehat{\beta}_1 W(s_0)^{1/3}\}\{\sigma^2(s_0) - 2(\boldsymbol{\lambda}^{\mathrm{T}}\gamma_0)\}$$

where $\boldsymbol{\lambda}$ is the vector of kriging weights as presented in (2.4), γ_0 is the vector of variogram values between s_0 and observed locations, and $\sigma^2(s_0)$ is the kriging variance.

Proof: see the Appendix A.

The assumption made on $\widehat{\beta}_0$ and $\widehat{\beta}_1$ is reasonable since we pooled the seasonal data to estimate the intercept and slope. The large amount of seasonal data justifies our assumption that $\widehat{\beta}_0$ and $\widehat{\beta}_1$ can be treated as fixed.

We predict rainfall at the 46 stations for the 143 days as discussed in Section 2.3.2, and calculate the statistics as defined in Section 2.2.4. Note $L = 143$ here. The results are: $D = 1344.852$mm, EB $= 7.044308$, EE $= 0.8534031$, and SSPE $= 332543.4\text{mm}^2$. In contrast, $\widetilde{D} = 7557.22$mm, $\widetilde{\text{EB}} = 39.58457$, $\widetilde{\text{EE}} = 0.6013614$, and $\widetilde{\text{SSPE}} = 374383.9\text{mm}^2$. By comparing D to $\widetilde{D}$, EB to $\widetilde{\text{EB}}$, and EE to $\widetilde{\text{EE}}$, we can see our prediction has a significantly less total difference in precipitation, lower estimation bias and higher estimation efficiency. This result indicates that our prediction does much better than the NexRad in estimating the total rainfall.

To evaluate our method in doing point estimation, we Compare SSPE with $\widetilde{\text{SSPE}}$, which shows that our prediction improves by approximate 11.2% over the nearest NexRad. September 11th is the highest rainfall day in 2003. The total rainfall amount of the 60 stations on that day is 1917.7mm, which is about twice the next largest rainfall amount of 1092.5mm on February 20th. Our prediction method does not work well on this "outlier" day. If we remove September 11th, the results are as following: $\text{SSPE}' = 306418.6\text{mm}^2$, and $\widetilde{\text{SSPE}}' = 367913\text{mm}^2$. Our prediction improves 16.7% compared to using the closest NexRad as the predictor.

For further evaluation of our prediction in terms of point estimation, we also

calculate the mean(and variance) of $\widehat{Z}_{ij} - Z_{ij}$, which is 0.2039200mm (50.38962mm^2), and the mean(and variance) of $W_{ij} - Z_{ij}$, which is 1.145901mm (55.46316mm^2). $\widehat{Z}_{ij}$, Z_{ij}, and W_{ij} are defined as in Section 2.2.4. It shows our prediction is less biased and slightly sharper than the closest NexRad to predict the true rainfall.

Figures 11 and 12 show the comparison of predictions on April 23rd and July 6th as examples to illustrate how the predictions behave. The side by side box plot of differences between prediction and the station data indicates that the predictions using our method are better than using the nearest NexRad location in the sense of both unbiasedness and low variability.

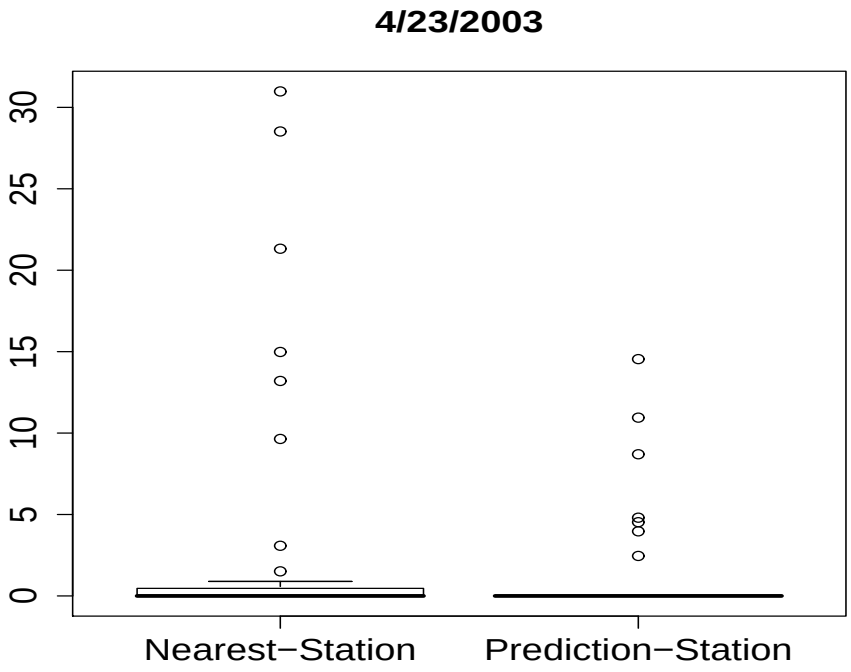

Figure 11: A comparison between predictors on 4/23/2003

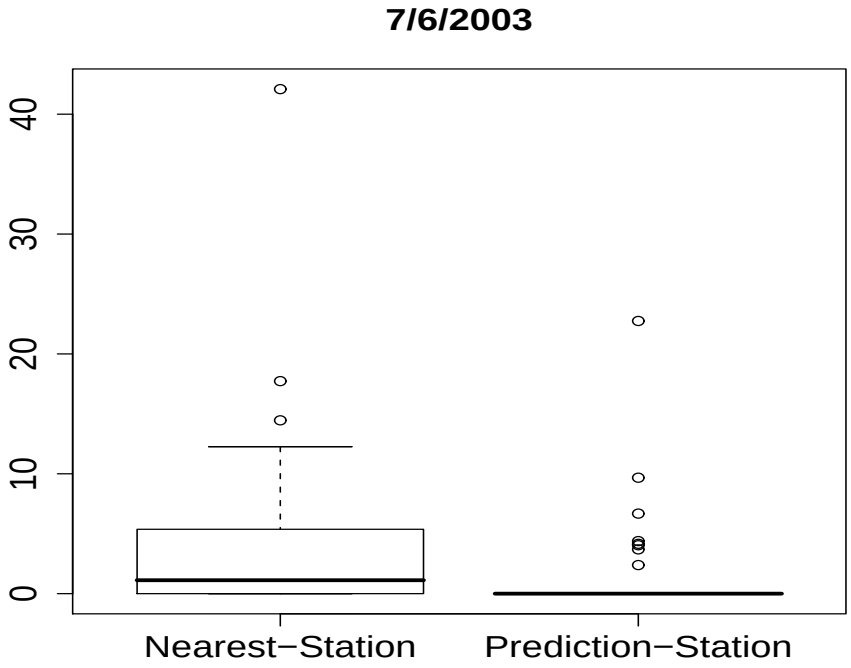

Figure 12: A comparison between predictors on 7/6/2003

2.4 A Simulation

2.4.1 Setup

To verify our procedures in the analysis, we carry out simulations to check their validity. Our goal is to make the simulation as comparable as possible to the actual data to judge the reliability of our improvement over using the closest NexRad. We simulate random fields on a 224 $\times$ 224 grid with a grid interval equal to 1. Each simulated random field contains 50,176 grid points, which corresponds to 50,151 NexRad grid points. Since the actual NexRad grid interval is 4, we shrink all the distance parameters in the actual data by the approximate ratio of 1/4 to get the corresponding ones in the simulation. The set up of the simulation is as follows:

2.4.1.1 Generating Rainfall

We generate NexRad data, $W^{1/3}$, as a random field on the 224 $\times$ 224 grid with an exponential covariance structure: $(\mu = 0.5, \sigma^2 = 2.0, \phi = 13)$. The value of σ^2 reflects the real data, and $\phi = 13$ corresponds to $\phi = 52$km in the real data. The value of μ is selected to make the ratio of positive values in the random field close to the actual rainfall data.

We set $\beta_0 = -0.3143$ and $\beta_1 = 0.938$, which are exactly the same as estimators from the cold season.

We generate spatial errors, ϵ, as a Gaussian random field on the 224 $\times$ 224 grid with mean zero and Matérn covariance (nugget = 0, sill = 1, range = 4.2, smoothness = 0.84). These values correspond to the median of estimators from the 2003 NexRad data.

We generate the corresponding gauge data, X, as:

$$X^{1/3} = [\beta_0 + \beta_1 W^{1/3} + \epsilon]^+$$

$[\cdot]^+$ denotes the clipped random field of $\cdot$ with zero as the threshold. We generate a clipped random field as in De Oliveira (2000). Specifically we generate the NexRad and gauge data over the same grid. The NexRad is the observed data, while the gauge data is considered the target data. Only a small part of the gauge data is assumed to be observed, and our goal is to predict the unobserved gauges using the observed gauges and the NexRad Data.

2.4.1.2 Parameter Estimation

We randomly pick 311 grid points from the 224 $\times$ 224 grid, and assume we observe the gauge data at these locations. Thus we have 311 pairs of observed (X, W) which correspond to the 311 gauges in Section 2.3.

Completely analogous to our procedures in Section 2.3 we pool the seasonal data to estimate β_2, β_0 and β_1. Here we generate 40 NexRad random fields and their corresponding gauge random fields, pool up all the $40 \times 311 = 12{,}440$ pairs of observed (X, W) to estimate β_2, and then β_0 and β_1. Only 7,285 of 12,440 pairs have $W > 0$ and are actually used in the estimation, which is comparable to the 7,114 pairs being used in the cold season.

2.4.1.3 Prediction

Again we randomly select 311 grid points as the observed gauge locations to parallel the 311 gauges in the data. Fix these locations in this step.

We find the nearest 9 NexRad W_{ij}, j=1,... 9 for each selected gauge X_i, which corresponds to the nearest 8 NexRad locations in the true data. In the real data, we assume that the gauge and it's nearest 8 NexRad have the same true rainfall. Likewise in the simulation we assume the gauge and it's nearest 9 NexRad have the same true rainfall. We then calculate residuals at each of the nearest 9 NexRad locations.

We employ the approximate likelihood method as in Section 2.3.2 to estimate the parameters in the Matérn covariance function. In the data analysis, the moving window is a circle of radius 50km centered on the gauge. This window contains about 500 NexRad locations, so in the simulation we set the moving window as a 23×23 grid centered on the gauge which contains 529 NexRad grid points.

Finally, we randomly pick 4000 grid points and predict the rainfall, $\widehat{Z}_i$, $i = 1, ..., 4000$, at those locations, and compare the prediction to the "true" rainfall Z_i.

We repeat subsections 2.4.1.1 to 2.4.1.3 200 times. We compute EB, EE, and $\widetilde{\text{EB}}$, $\widetilde{\text{EE}}$ which are defined as in Section 2.2.4. The only change we make in the definition in Section 2.2.4 is replacing 46 by 4000 and setting $L = 200$. For each simulation run, we define $\text{SSPE} = \sum_{i=1}^{4000}(\widehat{Z}_i - Z_i)^2$ for our prediction and $\widetilde{\text{SSPE}} = \sum_{i=1}^{4000}(W_i - Z_i)^2$ for the

nearest NexRad which is collocated with its corresponding gauge in the simulation. We did not calculate D and $\widetilde{D}$ in the simulation for two reasons. The value of D here is not comparable with the corresponding value in Section 2.3.3 due to a different number of locations at which the simulation makes predictions than the actual data analysis does, and the values of EB and $\widetilde{\text{EB}}$ are sufficient to represent the relative relationship between D and $\widetilde{D}$.

2.4.2 Simulation results and comparison with data analysis

The results of the simulation are: EB = -4.17782, EE = 0.9485363, and $\widetilde{\text{EB}}$ = -2.210882, $\widetilde{\text{EE}}$ = 0.9325573. The mean(and std. dev.) of SSPE over the 200 simulations is 216373.2 (8968.972). The mean(and std. dev.) of $\widetilde{\text{SSPE}}$ over the 200 simulations is 280582.6 (11477.85).

The simulated NexRad and gauge random field already agree in their total rainfall, so there is no room for improvement. However, comparison of SSPE with $\widetilde{\text{SSPE}}$ shows a 22.88% improvement of our prediction over the NexRad. The simulation indicates greater improvement than seen in the actual data analysis. This is not surprising as we have generated the simulated rainfall according to the model. Nevertheless the improvement here is roughly comparable to that seen in the data.

We show one simulated NexRad random field with its corresponding gauge and prediction random field in Figures 13, 14 and 15. The blank regions in those plots represent dry regions. Except the blank regions, the color scheme is shown in the plot. It is seen that the prediction follows the NexRad in discriminating wet areas from dry areas, but if we take a closer look at the light areas within the wet regions, we see that the pattern of highly lit areas in the prediction random field is closer to that in the gauge random field than is the NexRad random field. This indicates that we capture the essential areas of rainfall by successfully calibrating at the areas with

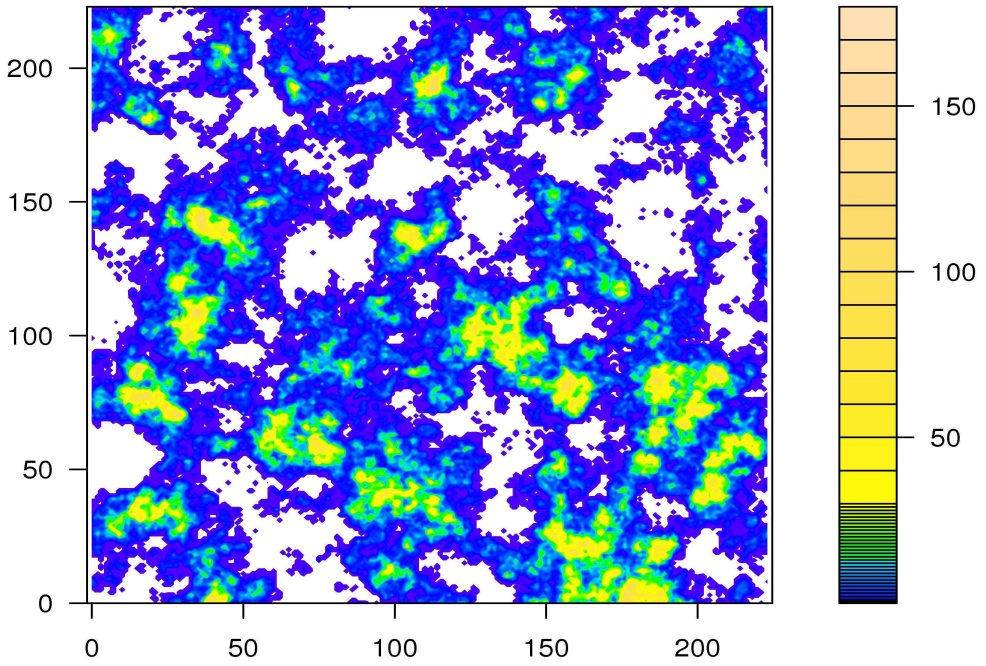

Figure 13: Simulated NexRad random field over 224×224 grid

large rainfall.

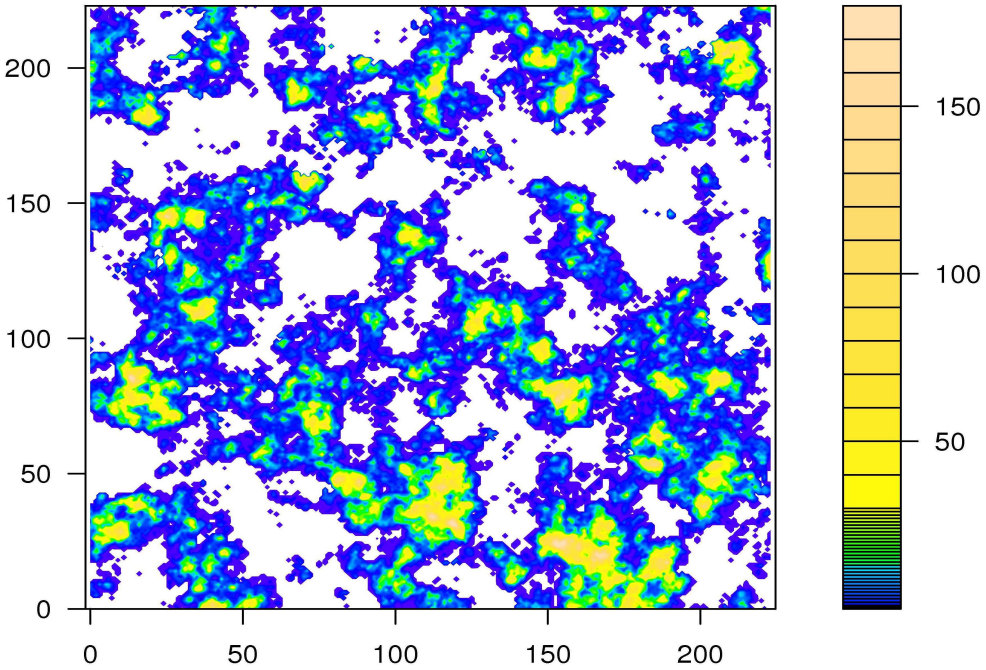

Figure 14: Simulated gauge random field

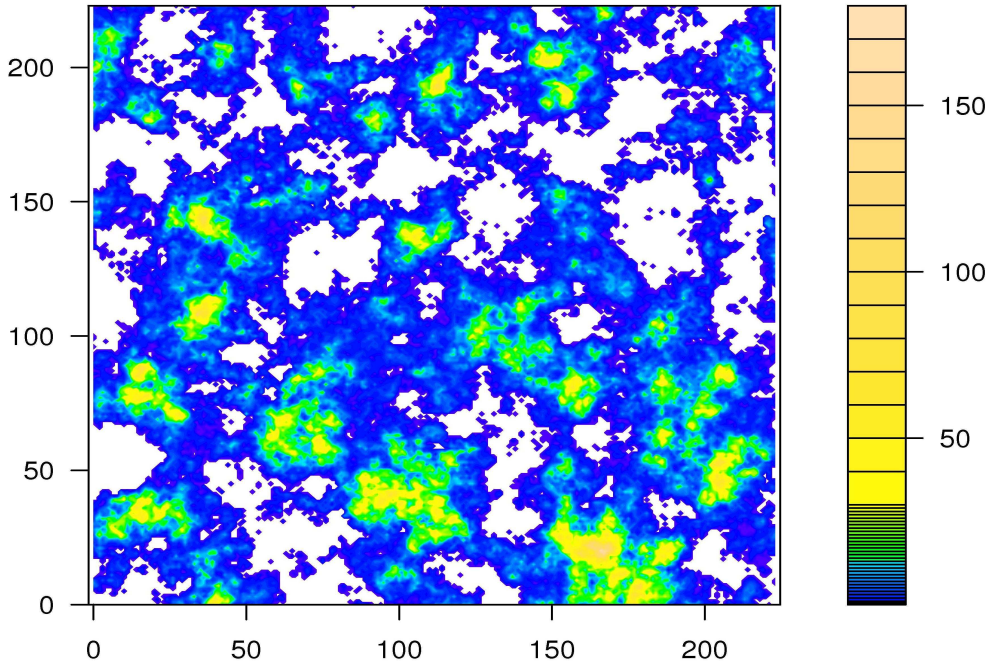

Figure 15: Prediction of gauge random field

2.5 Discussion

Hydrology data usually turns out to be of large size. The cumbersome body of data and the consequent massive work in manipulation increase the difficulty for data analysis. This paper proposed a simple, but efficient method to calibrate the NexRad data using the rain gauge data. The proposed method combines spatial rainfall occurrence estimation, bias reduction by regression techniques and geostatistical procedures. An approximate likelihood method is employed to estimate the parameters in the random process. Although the proposed method is developed based on 2003 Texas rainfall data, it provides a framework to analyze short time accumulated rainfall so that only slight modifications are needed to accommodate different data sets. For example, if the variance of the errors depends on the mean value, that is, the estimated $\xi(W, \beta)$ in (2.1) is some function of $\beta_0 + \beta_1 W^{1/k}$, we can employ weighted least squares to estimate β_0, β_1 and $\xi(W, \beta)$ iteratively.

The superiority of the proposed method lies in estimating total rainfall as well as point rainfall amount. The total rainfall is crucial in predicting flood, designing sewage system and managing other water resource decision support systems. Compared to the NexRad, our prediction of total rainfall amount has closer agreement with the true value. The estimation bias, estimation efficiency and root mean square difference are all highly improved over the NexRad itself. This method can be looked as a benchmark to calibrate the NexRad rainfall data using the gauge data. It can also be used in other environmental problems with a common structure.

Unlike the procedures in Barancourt et al. (1992), our method is applied to daily data and the geostatistical method is applied to the regression residuals. Thus there is no need to be concerned about nonergodicity and nonstationarity in the rainfall data.

We estimate the varying coefficients $\beta_2(t)$, $\beta_0(t)$ and $\beta_1(t)$ by fitting a simple linear model at the neighborhood of each t_0, however, the noise and the specific pattern of the varying coefficients motivates us to finally estimate the parameters from the seasonal data. Nevertheless, we are aware that other data sets may not exhibit such a specific pattern to allow us to consider the parameters as constant over a long time period. In this situation, we need an enhancement to the plane fitting method to refine the varying coefficients estimates. The approach of Cleveland, Gross, and Shyu (1991) assigns observations weights from a tricube weight function based on their Euclidian distance from t_0 in addition to the restriction to neighborhood. Specifically, the closer to t_0, the bigger the weight. In Cleveland's approach, varying coefficients are estimated by fitting a linear model using weighted least squares. By choosing the appropriate neighborhood size, this produces an adequate smooth and stable varying coefficients estimates, which makes the pointwise estimates safe to use.

Alternatively, spatiotemporal hierarchical Bayesian modeling (e.g., Wikle, Mil-

liff, Nychka, and Berliner, 2001) can be an effective method to combine data from different sources. Their strategy simplifies the complex problem by formulating three primary statistical models or stages. Stage 1 models only measurement errors, stage 2 formulates the true process and stage 3 specifies the priors. To implement this method, the computation is of the utmost concern. For huge data sets, like the NexRad and gauge data set being used here, this Bayesian approach is not easily carried out at least on the regular PC.

The model in this article does not account for the situation when the gauge is positive but the NexRad is below the threshold. This is not a big concern for the data set being used here, because on one hand, our estimated thresholds for both seasons are low and consequently only approximately 4% of the data fall into this situation. On the other hand, most of the positive gauge measurements among these 4% are of small values which usually are not of interest in practice. However, in the future, more research is needed to address this issue. Another very natural and interesting topic for future research is: how to make predictions using the NexRad estimates and rain gauge measurements? This question requires a deeper and further consideration of the spatio-temporal process of the rainfall.

CHAPTER III

ON THE ASYMPTOTIC JOINT DISTRIBUTION OF SAMPLE SPACE-TIME COVARIANCE ESTIMATORS

3.1 Introduction

Let $\{Z(\mathbf{x}) : \mathbf{x} \in D \subset \mathbb{R}^d\}$, $d \geq 1$, be a strictly stationary random field with covariance function

$$C(\mathbf{k}) = \text{cov}\{Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{k})\},$$

where $\mathbf{k}$ denotes an arbitrary lag in $\mathbb{R}^d$. Let $\mathbf{\Lambda}$ be a set of lags and let m denote the cardinality of $\mathbf{\Lambda}$. Define $\mathbf{G} = \{C(\mathbf{k}) : \mathbf{k} \in \mathbf{\Lambda}\}$ to be the length m vector of covariances at lags in $\mathbf{\Lambda}$. Let $\widehat{C}_n(\mathbf{k})$ denote the sample estimator of $C(\mathbf{k})$ based on observations in a sequence of increasing index sets $D_n \subset D$ and $\widehat{\mathbf{G}}_n = \{\widehat{C}_n(\mathbf{k}) : \mathbf{k} \in \mathbf{\Lambda}\}$ denote the estimator of $\mathbf{G}$. We are interested in the asymptotic distribution of $\widehat{\mathbf{G}}_n$. This distribution is important for several reasons such as testing directional properties of a random field. For example, Guan, Sherman, and Calvin (2004) and Lu and Zimmerman (2001) derived the asymptotic distribution of the spatial variogram when $d = 2$. Surprisingly, the distributional property of sample covariances has been investigated only in particular situations in the literature. Under the assumption of a Poisson process for modeling the observations' locations, Masry (1983) proved the asymptotic joint normality of sample autocovariances for time series and Karr (1986) generalized this result to a random field. Brockwell and Davis (1991, p. 229) and Fuller (1996, p. 333) derived the asymptotic joint normality of sample autocovariances under mild assumptions for stationary time series. Recognizing various data structures in practice, we consider several situations in a random field depending

on whether the observations are regularly spaced or irregularly spaced, and whether one part or the whole domain of interest is fixed or increasing. We also allow one dimension to denote time. We make no assumptions on the marginal or the joint distribution of observations other than mild moment and mixing conditions on the random field.

To formally state the asymptotic properties of the vector of sample covariances $\widehat{\mathbf{G}}_n$, we need to quantify the strength of dependence in the random field taking account of different types of spacing of observations. Following Rosenblatt (1956), we define the strong mixing coefficients for a random field with regularly spaced observations as:

$$\alpha_b(r) = \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathfrak{F}(E_i), |E_i| \leq b, i = 1, 2, d(E_1, E_2) \geq r\}, \tag{3.1}$$

where $|E|$ denotes the cardinality of the set E, $\mathfrak{F}(E)$ denotes the σ-algebra generated by the random variables $\{Z(\mathbf{x}) : \mathbf{x} \in E\}$, and $d(E_1, E_2) = \inf\{\sup_j |\mathbf{x}_{1j} - \mathbf{x}_{2j}| : \mathbf{x}_1 \in E_1, \mathbf{x}_2 \in E_2, j = 1, ..., d\}$. The supremum in (3.1) is taken over all compact and convex subsets $E_1 \subset \mathbb{R}^d$ and $E_2 \subset \mathbb{R}^d$ such that $d(E_1, E_2) \geq r$. For a random field with irregularly spaced observations, we define the mixing coefficients following Politis, Paparoditis, and Romano (1998). This definition, denoted by (3.1′), is formed by imposing an additional condition that E_2 is a shift of E_1 in (3.1). Note that $|E|$ denotes the Lebesgue measure (volume) of E in (3.1′).

If the observations are independent, then $\alpha_b(r) = 0$ for all $r > 0$. We need $\alpha_b(r)$ to approach 0 for large r, at some rate depending on b. Specifically, we decompose D_n into $D_n = \mathcal{F} \times \mathcal{I}_n$, where $\mathcal{F} \subset \mathbb{R}^p$, $\mathcal{I}_n \subset \mathbb{R}^q$ and $p + q = d$. Suppose $\mathcal{F}$ is a fixed space in the sense that finitely many observations are located within this space, and $\mathcal{I}_n$ is an increasing space. Following Sherman and Carlstein (1994), we assume the

mixing condition

$$\sup_b \frac{\alpha_b(r)}{b} = \mathrm{O}(r^{-\epsilon}) \text{ for some } \epsilon > q. \tag{C1}$$

We account for the shape of the domain in which we observe data by assuming:

$$|\mathcal{I}_n| = \mathrm{O}(n^q), \quad |\partial\mathcal{I}_n| = \mathrm{O}(n^{q-1}), \tag{C2}$$

where $\partial\mathcal{I}_n$ denotes the boundary of $\mathcal{I}_n$. This allows for a wide variety of domains. For example, let $\mathcal{A}$ denote the interior of a closed hypersurface contained in a q-dimensional hypercube with edge length 1. Denote by $\mathcal{A}_n$ the inflation of $\mathcal{A}$ by a factor n. Then $\mathcal{A}_n$ satisfies (C2) (Sherman, 1996).

If a random field provides sufficient information so that we can estimate its covariance at any lag smaller than a certain value, this random field is said to exhibit continuous lags. We then use a kernel estimator to estimate covariances for continuous lags. For example, if the observations are irregularly spaced in the subspace $\mathbb{R}^{q_1}$ of $\mathbb{R}^q$ ($q_1 \leq q$), covariances for any lag value in $\mathbb{R}^{q_1}$ can be estimated by a kernel estimator, see Section 3.3. However, if covariances at only discrete lags are estimable, this random field is said to exhibit discrete lags, and we usually use a moment estimator to estimate the covariances in this situation.

In what follows, we assume the mean of Z is known and equal to 0. If we remove this assumption, let $\widehat{C}_n^*(\mathbf{k})$ and $\widehat{\mathbf{G}}_n^*$ denote the mean corrected estimators of $C(\mathbf{k})$ and $\mathbf{G}$, respectively. We show in Lemma B.1.7 that $\widehat{\mathbf{G}}_n^*$ and $\widehat{\mathbf{G}}_n$ have the same asymptotic properties. Let $\widehat{C}_n(\mathbf{k})$ denote a moment estimator or a kernel estimator of the covariance $C(\mathbf{k})$ under the zero mean assumption and let λ_n be the bandwidth of the kernel estimator. We assume the moment condition for the moment estimator:

$$\sup_n \mathrm{E}\{|\sqrt{|\mathcal{I}_n|}\{\widehat{C}_n(\mathbf{k}) - C(\mathbf{k})\}|^{2+\delta}\} \leq C_\delta \text{ for some } \delta > 0, C_\delta < \infty. \tag{C3}$$

The moment condition (C3′) for the kernel estimator can be obtained from (C3) simply by replacing $|\mathcal{I}_n|$ with $|\mathcal{I}_n|\lambda_n^{q_1}$ and $C(\mathbf{k})$ with $\mathrm{E}\{\widehat{C}_n(\mathbf{k})\}$. The moment condition

is only slightly stronger than the existence of the (standardized) asymptotic variance of $\widehat{C}_n(\mathbf{k})$.

In this article, we derive the asymptotic joint distribution of sample covariances for space-time random fields $\{Z(\mathbf{s},t) : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}\}$. However, the results for this $\mathbb{R}^2 \times \mathbb{R}$ space can be easily extended to the $\mathbb{R}^d$ $(d > 3)$ space. Note that we consider an increasing domain asymptotic framework. For a recent comparison of infill and increasing domain asymptotics and the consequences for maximum likelihood estimators of covariance parameters between the two, see Zhang and Zimmerman (2005).

The rest of the paper is organized as follows. Section 2 demonstrates the asymptotic joint normality of sample space-time covariances for discrete lags, while Section 3 addresses the distributional behavior in the case of continuous lags. These two sections consider space-time data structure which are common in applications. Section 4 presents a simulation experiment. Appendix A states and proves some useful lemmas. Appendix B contains proofs of all theorems.

3.2 Discrete Lags

3.2.1 Regularly spaced observations with an increasing spatio-temporal domain

Consider a strictly stationary spatio-temporal random field $\{Z(\mathbf{s},t) : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}\}$. Let $D_n = S_n \times T_n$ be a finite set of lattice points in $\mathbb{Z}^2 \times \mathbb{Z}$ at which observations are taken. We allow the lattice in S_n and T_n to be defined in different metrics. Let $\mathbf{h}$ denote a lag in space and u denote a lag in time. The classical estimator of the covariance, i.e., the sample covariance, is given by

$$\widehat{C}(\mathbf{h},u) = \frac{1}{|D_n(\mathbf{h},u)|} \sum_{D_n(\mathbf{h},u)} Z(\mathbf{s},t)Z(\mathbf{s}+\mathbf{h},t+u),$$

where the sum is over $D_n(\mathbf{h}, u) = \{(\mathbf{s}, t) : (\mathbf{s}, t) \in D_n, (\mathbf{s} + \mathbf{h}, t + u) \in D_n\}$ and $|D_n(\mathbf{h}, u)|$ is the number of distinct elements in $D_n(\mathbf{h}, u)$.

We define the strong mixing coefficients of this random field as in (3.1), and assume the mixing condition, boundary condition and moment condition as in (C1), (C2) and (C3) setting $p = 0$, $q = 3$ and $\mathcal{I}_n = D_n$.

Theorem 3.2.1. *Let $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}\}$ be a strictly stationary spatio-temporal random field observed at lattice points in $D_n \subset \mathbb{Z}^3$ satisfying condition* (C2)*. Assume*

$$\sum_{\mathbf{s}\in\mathbb{Z}^2} \sum_{t\in\mathbb{Z}} |\mathrm{cov}\{Z(\mathbf{0}, 0)Z(\mathbf{h}_1, u_1), Z(\mathbf{s}, t)Z(\mathbf{s} + \mathbf{h}_2, t + u_2)\}| < \infty, \tag{3.2}$$

for all finite $\mathbf{h}_1, \mathbf{h}_2, u_1$ and u_2. Then $\Sigma = \lim_{n\to\infty} |D_n|\mathrm{cov}(\widehat{\mathbf{G}}_n, \widehat{\mathbf{G}}_n)$ exists, the (i, j)-th element of which is

$$\sum_{\mathbf{s}\in\mathbb{Z}^2} \sum_{t\in\mathbb{Z}} \mathrm{cov}\{Z(\mathbf{0}, 0)Z(\mathbf{h}_i, u_i), Z(\mathbf{s}, t)Z(\mathbf{s} + \mathbf{h}_j, t + u_j)\}.$$

If we further assume that Σ is positive definite and that conditions (C1) *and* (C3) *hold, then $\sqrt{|D_n|}(\widehat{\mathbf{G}}_n - \mathbf{G}) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, \Sigma)$.*

3.2.2 Observations with a fixed spatial domain and an increasing temporal domain

In many situations, the observations are taken from a fixed space $S \subset \mathbb{R}^2$ at regularly spaced times T_n. Let $S(\mathbf{h}) = \{\mathbf{s} : \mathbf{s} \in S, \mathbf{s} + \mathbf{h} \in S\}$ and $|S(\mathbf{h})|$ be the number of elements in $S(\mathbf{h})$.

In this particular case, we define the mixing coefficient (e.g., Ibragimov and Linnik, 1971, p. 306) only in the time direction, since the domain is increasing only in this direction:

$$\alpha(u) = \sup_{A,B}\{|P(A \cap B) - P(A)P(B)|, A \in \mathfrak{F}_{-\infty}^0, B \in \mathfrak{F}_u^\infty\},$$

where $\mathfrak{F}_{-\infty}^0$ is the σ-algebra generated by the past time process until $t = 0$, and $\mathfrak{F}_u^\infty$ is the σ-algebra generated by the future time process from $t = u$. The mixing coefficient

$\alpha(u)$ satisfies the strong mixing condition:

$$\alpha(u) = \mathrm{O}(u^{-\epsilon}) \text{ for some } \epsilon > 0, \tag{3.3}$$

and we also observe:

$$T_n = \{1, ..., n\}, \quad |\partial T_n| = 2 = \mathrm{O}(1). \tag{3.4}$$

Thus, conditions (3.3) and (3.4) are simply special cases of (C1) and (C2), respectively, by setting $p = 2$, $q = 1$, $b = \infty$, $\mathcal{F} = S$ and $\mathcal{I}_n = T_n$. We further assume the moment condition as in (C3) and define the estimator of C as:

$$\widehat{C}(\mathbf{h}, u) = \frac{1}{|S(\mathbf{h})||T_n|} \sum_{S(\mathbf{h})} \sum_{t=1}^{n-u} Z(\mathbf{s}, t) Z(\mathbf{s} + \mathbf{h}, t + u).$$

Corollary 3.2.2. *Let $\{Z(\mathbf{s}, t), \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}\}$ be a strictly stationary spatio-temporal random field observed in $D_n = S \times T_n$, where $S \subset \mathbb{R}^2$ and T_n satisfies condition (3.4). Assume*

$$\sum_{t \in \mathbb{Z}} |\mathrm{cov}\{Z(\mathbf{0}, 0) Z(\mathbf{h}_1, u_1), Z(\mathbf{s}, t) Z(\mathbf{s} + \mathbf{h}_2, t + u_2)\}| < \infty, \tag{3.5}$$

for all finite $\mathbf{h}_1, \mathbf{h}_2, u_1, u_2$ and $\mathbf{s} \in S$. Then $\Sigma = \lim_{n \to \infty} |T_n| \mathrm{cov}(\widehat{\mathbf{G}}_n, \widehat{\mathbf{G}}_n)$ exists, the (i, j)-th element of which is

$$\frac{1}{|S(\mathbf{h}_i)||S(\mathbf{h}_j)|} \sum_{S(\mathbf{h}_i)} \sum_{S(\mathbf{h}_j)} \sum_{t \in \mathbb{Z}} \mathrm{cov}\{Z(\mathbf{s}_1, 0) Z(\mathbf{s}_1 + \mathbf{h}_i, u_i), Z(\mathbf{s}_2, t) Z(\mathbf{s}_2 + \mathbf{h}_j, t + u_j)\}.$$

If we further assume that Σ is positive definite and that conditions (3.3) and (C3) *hold, then $\sqrt{|T_n|}(\widehat{\mathbf{G}}_n - \mathbf{G}) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, \Sigma)$.*

In Corollary 3.2.2, we allow the observations to be either regularly spaced or irregularly spaced in S. However, even for irregularly spaced observations, we only consider the covariances of observed spatial lags due to the limited number of observations in S. Note that in this section, we require the observations to be taken at

the same spatial locations over time, which is very common for monitoring stations. For example, the Irish wind data recently analyzed by Gneiting (2002), de Luna and Genton (2005) and Stein (2005) consists of time series of daily average wind speed at eleven meteorological stations in Ireland.

3.3 Continuous Lags

3.3.1 Spatially irregularly spaced observations with an increasing spatio-temporal domain

If the observations are spatially irregularly spaced in an increasing domain, we can then estimate the covariance for any spatial lag by employing kernel smoothing. Again consider a strictly stationary random field $\{Z(\mathbf{s},t) : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R}\}$. Let $D_n \subset \mathbb{R}^3$ denote the domain of interest in which observations are taken. We decompose D_n into $D_n = S_n \times T_n$, where S_n is an increasing index set, $S_n \subset \mathbb{R}^2$ and $T_n = \{1, ..., n\}$. We view the spatial locations at which Z is observed in S_n as *random* in number and location; specifically, they are generated from a homogeneous 2-dimensional Poisson process in $\mathbb{R}^2$ with intensity parameter ν.

Let N denote the random point process and $N(B)$ denote the random number of points of N contained in B, where B is any given Borel set. We further assume N to be independent of Z. We construct an estimator of covariance based on kernel smoothing.

Let $|S_n|$ denote the Lebesgue measure (not the cardinality) of S_n, and let $w(\cdot)$ be a bounded, symmetric density function on $\mathbb{R}^2$. Here and henceforth, we use $d\mathbf{s}$ to denote an infinitesimally small disc centered at $\mathbf{s}$. Define $N^{(2)}(d\mathbf{s}_1, d\mathbf{s}_2) \equiv N(d\mathbf{s}_1)N(d\mathbf{s}_2)I(\mathbf{s}_1 \neq \mathbf{s}_2)$, where $I(\mathbf{s}_1 \neq \mathbf{s}_2) = 1$ if $\mathbf{s}_1 \neq \mathbf{s}_2$ and 0 otherwise. The kernel covariance estimator over D_n is given by,

$$\widehat{C}_n(\mathbf{h}, u) = \frac{1}{\nu^2 |T_n||S_n|} \sum_{t=1}^{n-u} \int_{S_n} \int_{S_n} w_n(\mathbf{h} - \mathbf{s}_1 + \mathbf{s}_2) Z(\mathbf{s}_1, t) Z(\mathbf{s}_2, t+u) N^{(2)}(d\mathbf{s}_1, d\mathbf{s}_2).$$

where $w_n(\mathbf{x}) = \frac{1}{\lambda_n^2} w(\frac{\mathbf{x}}{\lambda_n})$, and λ_n is a sequence of positive constants satisfying condition (3.6) below. Here ν is assumed to be known, otherwise it can be consistently estimated by $\widehat{\nu} = \frac{N(S_n)}{|S_n|}$.

We define the mixing coefficients as in (3.1′) and assume the mixing and moment conditions as in (C1) and (C3′) setting $p = 0$, $q = 3$, $q_1 = 2$ and $\mathcal{I}_n = D_n$. In addition to the boundary condition in (C2), we need to consider the choice of bandwidth. Specifically, we assume

$$\lambda_n \to 0,\ \lambda_n^2 |S_n| \to \infty. \tag{3.6}$$

Define the fourth-order cumulant function (e.g., Karr, 1986):

$$\begin{aligned} Q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \ & \mathrm{E}\{Z(\mathbf{0})Z(\mathbf{x}_1)Z(\mathbf{x}_2)Z(\mathbf{x}_3)\} - C(\mathbf{x}_1)C(\mathbf{x}_3 - \mathbf{x}_2) \\ & - C(\mathbf{x}_2)C(\mathbf{x}_3 - \mathbf{x}_1) - C(\mathbf{x}_3)C(\mathbf{x}_2 - \mathbf{x}_1), \end{aligned} \tag{3.7}$$

where the $\mathbf{x}$'s denote generic locations, and $\mathbf{x} = (\mathbf{s}, t)$ in the spatio-temporal random field.

The following theorem states that $\widehat{C}_n(\mathbf{h}, u)$ is a consistent estimator for $C(\mathbf{h}, u)$ and $\widehat{\mathbf{G}}_n$ is asymptotically jointly normal under some conditions.

Theorem 3.3.1. *Let $\{Z(\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2,\ t \in \mathbb{R}\}$ be a strictly stationary spatio-temporal random field observed in $D_n = S_n \times T_n$, where $S_n \subset \mathbb{R}^2$ and $T_n = \{1, ..., n\}$. D_n satisfies condition* (C2). *Assume the locations $\mathbf{s}$ are generated by a homogeneous Poisson process in $\mathbb{R}^2$ and*

$$\sup_u \int_{\mathbb{R}^2} C(\mathbf{h}, u) d\mathbf{h} < \infty,$$

$$\sup_{\mathbf{h}, u, u'} \sum_{t \in \mathbb{Z}} \mathrm{E}\{Z(\mathbf{0}, 0)Z(\mathbf{h}, u)Z(\mathbf{0}, t)Z(\mathbf{h}, t + u')\} < \infty,$$

$$\sup_{t_1, t_2, t_3} \int_{\mathbf{s}_2 \in \mathbb{R}^2} |Q\{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), (\mathbf{s}_2 + \mathbf{s}_3, t_3)\}| d\mathbf{s}_2 < \infty,$$

for all $\mathbf{s}_1$ *and* $\mathbf{s}_3$. *Then* $\mathrm{E}\{\widehat{C}_n(\mathbf{h},u)\} \to C(\mathbf{h},u)$ *and* $\Sigma = \lim_{n\to\infty} |T_n||S_n|\lambda_n^2 \mathrm{cov}(\widehat{\mathbf{G}}_n, \widehat{\mathbf{G}}_n)$ *exists, the* (i,j)*-th element of which is*

$$\frac{1}{\nu^2}\int_{\mathbb{R}^2} w^2(\mathbf{x})d\mathbf{x}\sum_{t\in\mathbb{Z}} \mathrm{E}[Z(\mathbf{0},0)Z(\mathbf{h}_i,u_i)\{Z(\mathbf{0},t)Z(\mathbf{h}_i,t+u_j)I(\mathbf{h}_i=\mathbf{h}_j) \\ +Z(\mathbf{h}_i,t)Z(\mathbf{0},t+u_j)I(\mathbf{h}_i=-\mathbf{h}_j)\}],$$

where $I(\mathbf{h}_i = \pm\mathbf{h}_j) = 1$ *if* $\mathbf{h}_i = \pm\mathbf{h}_j$ *and 0 otherwise. If we further assume that* Σ *is positive definite and the conditions* (C1), (C3′) *and (3.6) hold, then*

$$\sqrt{n|S_n|}\lambda_n\{\widehat{\mathbf{G}}_n - \mathrm{E}(\widehat{\mathbf{G}}_n)\} \xrightarrow{d} \mathrm{N}_m(\mathbf{0},\Sigma).$$

3.3.2 Irregularly spaced observations with an increasing spatio-temporal domain

In Section 3.3.1, we discussed the properties of the kernel covariance estimator if the observations are irregularly spaced in S_n and regularly spaced in T_n. Another case occurs when the observations are irregularly spaced in the whole space $S_n \times T_n$. Karr (1986) showed mean-square consistency of the estimator $\widehat{C}$ of the covariance function C of the random field Z defined on $\mathbb{R}^d$ $(d \geq 2)$. It is an extension of the results in Masry (1983) who investigated the covariance estimator for time series, i.e., for random fields in $\mathbb{R}^1$.

In Karr's theorem, the random process, $\{Z(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^d\}$, is assumed to be a stationary Poisson process with intensity ν which is independent of the values of $Z(\mathbf{x})$. We specialize his results to our space-time random field by considering $\{Z(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^3\}$. Let $\mathbf{x}$ denote a location in $\mathbb{R}^3$, $d\mathbf{x}$ denote an infinitesimally small sphere centered at $\mathbf{x}$. Define $w(\cdot)$ and $N^{(2)}(d\mathbf{x}_1, d\mathbf{x}_2)$ in the same way as in Section 3.3.1 with the understanding that in this section the support of $w(\cdot)$ and $\mathbf{x}$ are defined in $\mathbb{R}^3$ rather than $\mathbb{R}^2$. The kernel estimator over D_n is defined as

$$\widehat{C}_n(\mathbf{k}) = \frac{1}{\nu^2|D_n|}\int_{D_n}\int_{D_n} w_n(\mathbf{k}-\mathbf{x}_1+\mathbf{x}_2)Z(\mathbf{x}_1)Z(\mathbf{x}_2)N^{(2)}(d(\mathbf{x}_1)d(\mathbf{x}_2)).$$

We define the mixing coefficients as in (3.1′) and assume the mixing condition, the boundary condition and the moment conditions (C1), (C2) and (C3′) setting $p = 0, q = 3, q_1 = 3$ and $\mathcal{I}_n = D_n$. We adopt the definition of Q from (3.7) in this section, and assume λ_n satisfies the condition:

$$\lambda_n \to 0,\ \lambda_n^3 |D_n| \to \infty. \tag{3.8}$$

Corollary 3.3.2. *Let $\{Z(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^3\}$ be a strictly stationary spatio-temporal random field which is observed in $D_n \subset \mathbb{R}^3$ satisfying condition* (C2)*. Assume the locations $\mathbf{x}$ are generated by a homogeneous Poisson process in $\mathbb{R}^3$. Assume $\int_{\mathbb{R}^3} C(\mathbf{k}) d\mathbf{k} < \infty$ and Q exists and satisfies*

$$\sup_{\mathbf{x}_1, \mathbf{x}_2} \int_{\mathbb{R}^3} |Q(\mathbf{x} + \mathbf{x}_1, \mathbf{x}, \mathbf{x}_2)| d\mathbf{x} < \infty.$$

Then $\mathrm{E}\{\widehat{C}(\mathbf{k})\} \to C(\mathbf{k})$ and $\Sigma = \lim_{n\to\infty} \lambda_n^3 |D_n| \mathrm{cov}(\widehat{\mathbf{G}}_n, \widehat{\mathbf{G}}_n)$ exists, the (i, j)-th element of which is

$$\frac{1}{\nu^2} \int_{\mathbb{R}^3} w^2(\mathbf{x}) d\mathbf{x} [\mathrm{E}\{Z^2(\mathbf{0}) Z^2(\mathbf{k}_i)\} I(\mathbf{k}_i = \pm \mathbf{k}_j)],$$

where $I(\mathbf{k}_i = \pm \mathbf{k}_j) = 1$ if $\mathbf{k}_i = \pm \mathbf{k}_j$ and 0 otherwise.

If we further assume $\mathrm{E}\{Z^2(\mathbf{0}) Z^2(\mathbf{k})\} > 0$ for all $\mathbf{k} \in \mathbf{\Lambda}$ and that conditions (C1),(C3′) *and (3.8) hold, then*

$$\sqrt{\lambda_n^3 |D_n|} \{\widehat{\mathbf{G}}_n - \mathrm{E}(\widehat{\mathbf{G}}_n)\} \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, \Sigma).$$

Unlike Theorem 3.3.1, Corollary 3.3.2 combines $\mathbf{s}$ and t as a single location $\mathbf{x}$. As a result, we get a more concise form for Σ than in Theorem 3.3.1. However, we can expect there will be two separate terms for $I(\mathbf{h}_i) = I(\mathbf{h}_j)$ and $I(\mathbf{h}_i) = I(-\mathbf{h}_j)$ in the formulation of Σ as in Theorem 3.3.1 if we consider $\mathbf{s}$ and t separately. The proof for asymptotic normality of $\widehat{\mathbf{G}}_n$ follows from Lemma B.1.5 setting $p = 3$, $q = 0$.

3.4 Simulation

In order to illustrate the approach to joint normality of $\widehat{\mathbf{G}}_n$ with the derived asymptotic covariance matrix, we perform the following simulation experiment. For simplicity, we consider only the situation given in Section 3.2.2, noting that it corresponds to many practical applications such as the Irish wind data mentioned at the end of Section 3.2.2. We simulate a random field via a vector autoregressive (VAR) model with spatial structure (de Luna and Genton, 2005). Specifically, we assume the fixed space S is a 3×3 grid with grid interval 1 and consider the VAR(1) model:

$$Z_t = RZ_{t-1} + \epsilon_t, \tag{3.9}$$

where $Z_t = (Z(\mathbf{s}_1, t), ..., Z(\mathbf{s}_9, t))^{\mathrm{T}}$, and ϵ_t is a Gaussian white noise process with a spatial stationary and isotropic exponential correlation function given by $(\Sigma_\epsilon)_{ij} = \exp(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\phi})$, $i, j = 1, ..., 9$, where ϕ denotes the range parameter (we set $\phi = 1$ initially). R is a 9×9 matrix of coefficients which determines the dependency between Z_t and $Z_{t-\tau}$ as $\mathrm{cov}(Z_t, Z_{t-\tau}) = R^\tau \Gamma_z(0)$, where $\mathrm{vec}\{\Gamma_z(0)\} = (I_{81} - R \otimes R)^{-1}\mathrm{vec}(\Sigma_\epsilon)$, and $\mathrm{vec}(\cdot)$ denotes the operator vectorizing a matrix. Note that the coefficients in R are determined only by $\mathbf{s}$, not by t. We set these coefficients as follows: for each $(\mathbf{s}_i, t)$, $i = 1, ..., 9$, the coefficient is 0.2 for $(\mathbf{s}_i, t-1)$, whereas it is 0.1 for $\{(\mathbf{s}_j, t-1) : \|\mathbf{s}_j - \mathbf{s}_i\| = 1\}$, $1 \leq j \leq 9$ and 0 for the remaining $(\mathbf{s}, t-1)$'s.

We choose two space-time lags, $\mathbf{k}_1 = (\|\mathbf{h}\| = 1, u = 0)$ and $\mathbf{k}_2 = (\|\mathbf{h}\| = 1, u = 1)$, and set $\mathbf{\Lambda} = \{\mathbf{k}_1, \mathbf{k}_2\}$. Here $\|\mathbf{h}\|$ denotes the Euclidian norm of $\mathbf{h}$. To explore the empirical joint distribution of $\widehat{\mathbf{G}}_n$ as $T_n = \{1, ..., n\}$ increases, we first set $n = 3$, simulate an $S \times T_n$ random field using (3.9) and compute $\widehat{C}(\mathbf{k}_1)$ and $\widehat{C}(\mathbf{k}_2)$ over this random field. We repeat this 5000 times obtaining a sample of $\widehat{C}(\mathbf{k}_1)$ and $\widehat{C}(\mathbf{k}_2)$ values. We then set $n = 10, 20, 50, 70, 100, 150, 200, 500, 1000, 5000$, respectively, and repeat the same procedure as with $n = 3$. We assess the joint normality of $\widehat{C}(\mathbf{k}_1)$ and

$\widehat{C}(\mathbf{k}_2)$ for each n using the multivariate measures of skewness and kurtosis introduced by Mardia (1970).

Finally, we evaluate the covariance matrix of $\widehat{\mathbf{G}}_n$. Denote by $\widehat{C}_{i,j}(u)$ the sample estimator of $C_{i,j}(u) = \text{cov}\{Z(\mathbf{s}_j, t), Z(\mathbf{s}_i, t+u)\}$. Priestley (1981, p. 693) presents a formula to compute $\text{cov}\{\widehat{C}_{i_1,j_1}(u_1), \widehat{C}_{i_2,j_2}(u_2)\}$ for large n when Z_t is a Gaussian process. The result of this is essentially an immediate form of Corollary 3.2.2 once we assume Z is a Gaussian process and we are interested only in $\widehat{C}_{i,j}(u)$ rather than $\widehat{C}(\mathbf{k})$. However, to obtain the theoretical values for the general $\widehat{C}(\mathbf{k})$ in Corollary 3.2.2 using Priestley's method, we need to find a link between the form in Corollary 3.2.2 and $\widehat{C}_{i,j}(u)$.

Lemma 3.4.1. *If* $Z_t = (Z(\mathbf{s}_1, t), ..., Z(\mathbf{s}_n, t))^{\mathrm{T}}$ *is a Gaussian process with mean 0,* $C_{i,j}(u) = E\{Z(\mathbf{s}_j, t)Z(\mathbf{s}_i, t+u)\}$ *and* $\widehat{C}_{i,j}(u) = \frac{1}{|T_n|}\sum_t Z(\mathbf{s}_j, t)Z(\mathbf{s}_i, t+u)$*, then for large* n*,*

$$\begin{aligned} & \text{cov}\{\widehat{C}_n(\mathbf{h}_i, u_i), \widehat{C}_n(\mathbf{h}_j, u_j)\} \\ = \; & \tfrac{1}{|S(\mathbf{h}_i)|}\tfrac{1}{|S(\mathbf{h}_j)|}\textstyle\sum_{\mathbf{s}_1 \in S(\mathbf{h}_i)}\sum_{\mathbf{s}_2 \in S(\mathbf{h}_j)} \text{cov}\{\widehat{C}_{\mathbf{s}_1+\mathbf{h}_i,\mathbf{s}_1}(u_i), \widehat{C}_{\mathbf{s}_2+\mathbf{h}_j,\mathbf{s}_2}(u_j)\}, \end{aligned}$$

where

$$\text{cov}\{\widehat{C}_{i_1,j_1}(s), \widehat{C}_{i_2,j_2}(u)\} \sim \frac{1}{|T_n|}\sum_{r\in\mathbb{Z}}\{C_{j_2,j_1}(r)C_{i_2,i_1}(r+u-s) + C_{i_2,j_1}(r+u)C_{j_2,i_1}(r-s)\}.$$

We calculate the theoretical values in Corollary 3.2.2 using Lemma 3.4.1, and compare them with the simulation output. The simulation output together with the theoretical values are summarized in Table 1. We see from this table that the multivariate measures of skewness and kurtosis approach 0 and 8 respectively as n increases. This agrees with the skewness and kurtosis of a bivariate normal. Additionally, all $|T_n|\text{cov}\{\widehat{C}(\mathbf{k}_1), \widehat{C}(\mathbf{k}_2)\}$, $|T_n|\text{var}\{\widehat{C}(\mathbf{k}_1)\}$ and $|T_n|\text{var}\{\widehat{C}(\mathbf{k}_2)\}$ approach their corresponding theoretical values as n increases. This verifies our asymptotic covariance matrix of $\widehat{\mathbf{G}}_n$. As seen in Table 1, $\text{cov}(\widehat{\mathbf{G}}_n)$ stabilizes at about $n = 100$.

Table 1: Simulation results for a 3×3 grid, $\phi = 1$

n	$b_{1,2}$	$b_{2,2}$	$\lvert T_n \rvert \text{cov}\{\widehat{C}(\mathbf{k}_1), \widehat{C}(\mathbf{k}_2)\}$	$\lvert T_n \rvert \text{var}\{\widehat{C}(\mathbf{k}_1)\}$	$\lvert T_n \rvert \text{var}\{\widehat{C}(\mathbf{k}_2)\}$
3	5.797	17.565	0.382	0.491	0.559
10	2.730	12.929	0.510	0.651	0.584
20	1.291	10.279	0.483	0.631	0.542
50	0.701	9.342	0.497	0.643	0.553
70	0.435	8.917	0.498	0.637	0.552
100	0.270	8.583	0.492	0.647	0.533
150	0.185	8.447	0.505	0.663	0.546
200	0.179	8.352	0.494	0.652	0.533
500	0.056	8.026	0.496	0.655	0.538
1000	0.043	8.084	0.511	0.665	0.554
5000	0.010	8.053	0.501	0.653	0.539
∞	0	8	0.497*	0.653*	0.539*

$b_{1,2}$ and $b_{2,2}$ denote multivariate measures of skewness and kurtosis, respectively.
* are obtained using Lemma 3.4.1.

We have observed that the simulation results do not change appreciably when the number of simulations exceeds 5000, so we show only the results based on 5000 simulations. Setting the range parameter ϕ in the spatial correlation function to be 1.5 increases the variances and covariances as in Table 2, but the trend of multivariate measures of skewness and kurtosis and the convergence of $\text{cov}(\widehat{\mathbf{G}}_n)$ remain the same as when $\phi = 1$. If we assume S is a 4×4 grid, the variances and covariances become smaller than with a 3×3 grid (see Table 3) as larger grid sizes provide more data to estimate $\widehat{\mathbf{G}}_n$. However, all our simulation experiments show a more rapid approach of $\widehat{\mathbf{G}}_n$ to joint normality as n increases.

Table 2: Simulation results for a 3×3 grid, $\phi = 1.5$

n	$b_{1,2}$	$b_{2,2}$	$\lvert T_n\rvert\text{cov}\{\widehat{C}(\mathbf{k}_1), \widehat{C}(\mathbf{k}_2)\}$	$\lvert T_n\rvert\text{var}\{\widehat{C}(\mathbf{k}_1)\}$	$\lvert T_n\rvert\text{var}\{\widehat{C}(\mathbf{k}_2)\}$
3	9.7449	25.7320	0.764	0.979	0.984
10	3.6833	15.0090	0.831	1.054	0.933
50	0.6318	8.9071	0.847	1.099	0.896
100	0.4137	8.7333	0.900	1.149	0.937
150	0.2186	8.3818	0.910	1.165	0.943
200	0.1909	8.4526	0.866	1.113	0.903
500	0.0998	7.8431	0.920	1.191	0.937
1000	0.0240	8.0432	0.927	1.199	0.938
∞	0	8	0.902^*	1.164^*	0.928^*

$b_{1,2}$ and $b_{2,2}$ denote multivariate measures of skewness and kurtosis, respectively.
* are obtained using Lemma 3.4.1.

Table 3: Simulation results for a 4×4 grid, $\phi = 1$

n	$b_{1,2}$	$b_{2,2}$	$\lvert T_n\rvert\text{cov}\{\widehat{C}(\mathbf{k}_1), \widehat{C}(\mathbf{k}_2)\}$	$\lvert T_n\rvert\text{var}\{\widehat{C}(\mathbf{k}_1)\}$	$\lvert T_n\rvert\text{var}\{\widehat{C}(\mathbf{k}_2)\}$
3	5.4151	17.9054	0.310	0.381	0.405
10	2.7941	13.3841	0.352	0.432	0.395
50	0.5359	8.8778	0.346	0.430	0.373
100	0.2685	8.6148	0.349	0.442	0.368
150	0.1561	8.2447	0.355	0.447	0.372
200	0.1206	8.3223	0.347	0.438	0.364
500	0.0430	8.0744	0.358	0.450	0.374
1000	0.0033	7.9753	0.346	0.435	0.366
∞	0	8	0.354^*	0.445^*	0.371^*

$b_{1,2}$ and $b_{2,2}$ denote multivariate measures of skewness and kurtosis, respectively.
* are obtained using Lemma 3.4.1.

CHAPTER IV

CONCLUSION AND FUTURE RESEARCH

4.1 Introduction

Let $\{Z(\mathbf{s},t) : \mathbf{s} \in \mathbb{R}^d, t \in \mathbb{R}\}$ be a strictly stationary spatio-temporal random process with covariance function $C(\mathbf{h},u) = \text{cov}\{Z(\mathbf{s},t), Z(\mathbf{s}+\mathbf{h},t+u)\}$, where $\mathbf{h}$ and u denote an arbitrary spatial lag and time lag, respectively. To make inferences from $Z(\mathbf{s},t)$, we often make assumptions regarding $C(\mathbf{h},u)$ which include full symmetry, separability, Taylor's hypothesis (Taylor, 1938) and isotropy as the most common ones. For example, $Z(\mathbf{s},t)$ has fully symmetric covariance if $C(\mathbf{h},u) = C(\mathbf{h},-u)$, or $C(\mathbf{h},u) = C(-\mathbf{h},u)$. Among the class of symmetric covariances, a covariance is separable if and only if $C(\mathbf{h},u) = \dfrac{C(\mathbf{h},0)C(\mathbf{0},u)}{C(\mathbf{0},0)}$, that is, the space-time covariance can be factored into the product of a purely spatial covariance and a purely temporal covariance. The relationship between several assumptions about the space-time process is illustrated in Gneiting, Genton, and Guttorp (2006). Taylor's hypothesis addresses the relationship between the purely spatial and the purely temporal covariance by examining if there exists a velocity vector $\mathbf{v} \in \mathbb{R}^d$ such that $C(\mathbf{0},u) = C(\mathbf{v}u,0)$. Spatial isotropy restricts $C(\mathbf{h},u)$ as a function dependent only on Euclidian norm of $\mathbf{h}$ and u, rather than the vector $\mathbf{h}$.

Although the assumptions for the covariances simplify model structures and ease the possible computational extension, they are not appropriate for many situations, especially in the study of geoscience, meterology and ecology, as the circulation of atmosphere can easily break the symmetry assumption in a space-time process. For example, Gneiting et al. (2006) found the lack of full symmetry and the lack of separability as well for the Irish wind data described in Haslett and Raftery (1989). Jun

and Stein (2004) modelled the sulfate concentration levels by a space-time asymmetric covariance function for the strong space-time asymmetry in the data. Cox and Isham (1988) discussed some restrictions for a specific covariance function to satisfy Taylor's hypothesis. Guan, Sherman, and Calvin (2004) provided evidence to reject the spatial isotropy assumption in a wind-speed data set and suggested that the correlation in the N-W direction is larger than in the E-W direction.

In order to check the validity of assumptions made for covariances, many approaches have been proposed to test one of those assumptions. Mitchell, Genton, and Gumpertz (2005) proposed a likelihood ratio test for separability of covariance models in the context of multivariate repeated measures, and Mitchell, Genton, and Gumpertz (2006) generalized this test to the spatio-temporal context. Scaccia and Martin (2005) presented a spectral method to test the symmetry and separability for spatial processes. Fuentes (2006) proposed a nonparametric test for spatial-temporal processes also based on a spectral method. Guan et al. (2004) developed a nonparametric test for spatial isotropy using subsampling based on the asymptotic joint distribution of sample variograms.

We provide a framework to assess all the assumptions for covariances discussed above, based on the asymptotic joint distribution of sample space-time covariance estimators derived in Chapter III. Most of our test statistics are linear combinations of sample covariance estimators. Depending on different hypothesis, we choose different linear combinations. To test for separability, we apply a log transform to the covariance estimators first and then obtain a linear combination of log transformations.

4.2 Various Assessments for Covariances

4.2.1 Test statistics

Let $\mathbf{\Lambda}$ be a set of lags and let m denote the cardinality of $\mathbf{\Lambda}$. According to the definitions of the assumptions, we give the null hypothesis for each test in the sequence of testing the full symmetry, separability, Taylor's hypothesis and spatial isotropy.

1. H_0: $C(\mathbf{h}, u) - C(\mathbf{h}, -u) = 0$, $(\mathbf{h}, u) \in \mathbf{\Lambda}$.

2. H_0: $\log\{C(\mathbf{h}, u)\} + \log\{C(\mathbf{0}, 0)\} - \log\{C(\mathbf{h}, 0)\} - \log\{C(\mathbf{0}, u)\} = 0$, $(\mathbf{h}, u) \in \mathbf{\Lambda}$.

3. H_0: $C(\mathbf{0}, u) - C(u\mathbf{v}, 0) = 0$ for some $\mathbf{v} \in \mathbb{R}^d$.

4. H_0: $C(\mathbf{h}_1, u) - C(\mathbf{h}_2, u) = 0$, $(\mathbf{h}_1, u), (\mathbf{h}_2, u) \in \mathbf{\Lambda}$, $\mathbf{h}_1 \neq \mathbf{h}_2$, but $\|\mathbf{h}_1\| = \|\mathbf{h}_2\|$.

Let D_n be the finite domain of observations, and $\widehat{C}_n(\mathbf{h}, u)$ denote the estimator of $C(\mathbf{h}, u)$ over D_n. Replacing $C(\cdot)$ with $\widehat{C}_n(\cdot)$ in the above four null hypotheses, we form the test statistic as the left hand side of the equation for each hypothesis.

TS1 $= \widehat{C}_n(\mathbf{h}, u) - \widehat{C}_n(\mathbf{h}, -u)$, $(\mathbf{h}, u) \in \mathbf{\Lambda}$.

TS2 $= \log\{\widehat{C}_n(\mathbf{h}, u)\} + \log\{\widehat{C}_n(\mathbf{0}, 0)\} - \log\{\widehat{C}_n(\mathbf{h}, 0)\} - \log\{\widehat{C}_n(\mathbf{0}, u)\}$, $(\mathbf{h}, u) \in \mathbf{\Lambda}$.

TS3 $= \widehat{C}_n(\mathbf{0}, u) - \widehat{C}_n(u\mathbf{v}, 0)$ for some $\mathbf{v} \in \mathbb{R}^d$.

TS4 $= \widehat{C}_n(\mathbf{h}_1, u) - \widehat{C}_n(\mathbf{h}_2, u)$, $(\mathbf{h}_1, u), (\mathbf{h}_2, u) \in \mathbf{\Lambda}$, $\mathbf{h}_1 \neq \mathbf{h}_2$, but $\|\mathbf{h}_1\| = \|\mathbf{h}_2\|$.

Let $\mathbf{G} = \{C(\mathbf{h}, u), (\mathbf{h}, u) \in \mathbf{\Lambda}\}$ and $\widehat{\mathbf{G}}_n = \{\widehat{C}_n(\mathbf{h}, u), (\mathbf{h}, u) \in \mathbf{\Lambda}\}$ denotes the estimator of $\mathbf{G}$ over D_n. Observe that TS1, TS3 and TS4 are linear combinations of the sample covariance estimators, and TS2 is a linear combination of the log transformation of the sample covariance estimators, thus TS1, TS3 and TS4 can be rewritten

into the form of $A_1\widehat{\mathbf{G}}_n$, $A_3\widehat{\mathbf{G}}_n$ and $A_4\widehat{\mathbf{G}}_n$, respectively, while TS2 can be rewritten into $A_2\log(\widehat{\mathbf{G}}_n)$ for a specified $\mathbf{\Lambda}$ and matrices A_1, A_2, A_3 and A_4. For example, if

$$\begin{aligned}\mathbf{\Lambda} = \; & \{(\mathbf{0},0),(\mathbf{h}_1,u_1),(\mathbf{h}_1,-u_1),(\mathbf{h}_2,u_2),(\mathbf{h}_2,-u_2),\\ & (\mathbf{h}_1,0),(\mathbf{0},u_1),(\mathbf{h}_2,0),(\mathbf{0},u_2),(u_1\mathbf{v},0),(u_2\mathbf{v},0)\},\end{aligned}$$

that is,

$$\begin{aligned}\mathbf{G} = \; & \{C(\mathbf{0},0),C(\mathbf{h}_1,u_1),C(\mathbf{h}_1,-u_1),C(\mathbf{h}_2,u_2),C(\mathbf{h}_2,-u_2),\\ & C(\mathbf{h}_1,0),C(0,u_1),C(\mathbf{h}_2,0),C(0,u_2),C(u_1\mathbf{v},0),C(u_2\mathbf{v},0)\},\end{aligned}$$

then

$$A_1 = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \end{bmatrix},$$

$$A_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}.$$

If we further assume $\mathbf{h}_1 \neq \mathbf{h}_2$, but $\|\mathbf{h}_1\| = \|\mathbf{h}_2\|$, $u_1 = u_2$,

$$A_4 = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

Under the null hypothesis, $A_1\mathbf{G} = A_3\mathbf{G} = A_4\mathbf{G} = A_2\log(\mathbf{G}) = \mathbf{0}$. It is easy to get the asymptotic distribution of the test statistics knowing the asymptotic joint normality of $\widehat{\mathbf{G}}_n$. Here we only consider the situation where the data are observed in a

fixed spatial domain S by an increasing temporal domain T_n as in Section 3.2.2, but all the other situations in 3.2.2 can be discussed in the same way. Since $\sqrt{|T_n|}(\widehat{\mathbf{G}}_n - \mathbf{G}) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, \Sigma)$, we obtain

$$\sqrt{|T_n|}\{\log(\widehat{\mathbf{G}}_n) - \log(\mathbf{G})\} \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, B^{\mathrm{T}}\Sigma B)$$

by the multivariate delta theorem (e.g., Mardia, Kent, and Bibby, 1979, p. 52), where B is a diagonal matrix with $B_{ii} = 1/\widehat{\mathbf{G}}_{ni}, i = 1, ..., m$.

We also know that under the null hypothesis,

$$|T_n|(A\widehat{\mathbf{G}}_n)^{\mathrm{T}}(A\Sigma A^{\mathrm{T}})^{-1}(A\widehat{\mathbf{G}}_n) \xrightarrow{d} \chi^2_p$$

and

$$|T_n|[A\{\log(\widehat{\mathbf{G}}_n)\}]^{\mathrm{T}}(AB^{\mathrm{T}}\Sigma BA^{\mathrm{T}})^{-1}[A\{\log(\widehat{\mathbf{G}}_n)\}] \xrightarrow{d} \chi^2_p$$

for a matrix A with row rank p. This provides the asymptotic distribution for all the test statics TS1, TS2, TS3 and TS4.

4.2.2 Choice of matrices A_1, A_2, A_3 and A_4

In Section 4.2.1, we can choose a subset of rows from A_4 to form a new matrix to test the spatial isotropy. For example, we pick only the first two rows, say, to get a new test, and the distribution of TS4 switches to χ^2_2 from χ^2_4. This is also applicable to the other matrices A_1, A_2 and A_3 for their corresponding tests. We may loose some power of the test by choosing only a subset of all the possible linear combinations, but if the size of $\mathbf{\Lambda}$ is large, we will make this choice for simplicity. However, we will deliberately rather than randomly choose a subset to retain the power. We prefer the smaller lags to the larger lags since usually the covariance estimators of smaller lags are obtained over more observations than larger lags. We also need to include a variety of lags in terms of both spatial and time to ensure that we are testing the

characteristic for the whole spatio-temporal process. To test the spatial isotropy, see Guan et al. (2004) for the details on how to make choices.

REFERENCES

Anagnostou, E., Krajewski, W., Seo, D., and Johnson, E. (1998). Mean-field rainfall bias studies for WSR-88D. *Journal of Hydrologic Engineering* **3**, 149–159.

Barancourt, C., Creutin, J. D., and Rivoirard, J. (1992). A method for delineating and estimating rainfall fields. *Water Resource Research* **28**, 1122–1144.

Bedient, P. B., Hoblit, B. C., Gladwell, D. C., and Vieux, B. E. (2000). NEXRAD radar for flood prediction in Houston. *Journal of Hydrologic Engineering* **5**, 269–277.

Bedient, P. B., Holder, A., Benavides, J. A., and Vieux, B. E. (2003). Radar-based flood warning system applied to tropical storm Allison. *Journal of Hydrologic Engineering* **8**, 308–318.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York: Chapman and Hall.

Cleveland, W. S., Gross, E., and Shyu, W. M. (1991). Local regression models. In J. M. Chambers and T. Hastie (eds.), *Statistical Models in S*, pp. 309–376. Pacific Grove, CA: Wadsworth and Brooks/Cole.

Cox, D. R. and Isham, V. (1988). A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London Ser. A* **415**, 317–328.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York: Wiley.

de Luna, X. and Genton, M. G. (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica* **15**, 547–568.

De Oliveira, V. (2000). Bayesian prediction of clipped gaussian random fileds. *Computational Statistics & Data Analysis* **34**, 299–314.

De Oliveira, V. (2004). A simple model for spatial rainfall fields. *Stochatic Environmental Research* **18**, 131–140.

Fuentes, M. (2006). Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning & Inference* **136**, 447–466.

Fuller, W. A. (1996). *Introduction to Statistical Times Series*. New York: Wiley.

Gel, Y., Raftery, E., and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method (with discussion). *Journal of the American Statistical Association* **99**, 575–587.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97**, 590–600.

Gneiting, T., Genton, M. G., and Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability and full symmetry. In B. Finkenstaedt, L. Held, and V. Isham (eds.), *Statistics of Spatio-Temporal Systems*, Monographs in Statistics and Applied Probability. Chapman & Hall/CRC Press, in press.

Guan, Y. (2003). *Nonparametric Methods of Assessing Spatial Isotropy*. Ph.D. thesis, Texas A&M University, Department of Statistics, College Station.

Guan, Y., Sherman, M., and Calvin, J. A. (2004). A nonparametric test for spatial isotropy using subsampling. *Journal of the American Statistical Association* **99**, 810–821.

Haas, T. C. (1995). Local prediction of a sptio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association* **85**, 950–963.

Handcock, M. S. and Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association* **89**, 368–390.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind-power resource (with discussion). *Applied Statistics* **38**, 1–50.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B* **55**, 757–796.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. New York: Springer.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika* **12**, 134–139.

Jayakrishnan, R., Srinivasan, R., and Arnold, J. (2004). Comparison of raingage and WSR-88D Stage III precipitation data over the Texas-Gulf basin. *Journal of Hydrology* **292**, 135–152.

Jun, M. and Stein, M. L. (2004). An approach to producing space-time covariance functions on spheres. Tech. rep. 18, CISES.

Karr, A. F. (1986). Inference for stationary random fields given poisson samples. *Advances in Applied Probability* **18**, 406–422.

Lu, H. and Zimmerman, D. L. (2001). Testing for isotropy and other directional symmetry properties of spatial correlation. *preprint. (Personal Collection, B. Li).*

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Masry, E. (1983). Non-parametric covariance estimation from irregularly-spaced data. *Advances in Applied Probability* **15**, 113–132.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology* **58**, 1246–1266.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2005). Testing for separability of space-time covariances. *Environmetrics* **16**, 819–831.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006). A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis* **97**, 1025–1043.

Politis, D. N., Paparoditis, E., and Romano, J. P. (1998). Large sample inference for irregularly spaced dependent observations based on subsampling. *Indian Journal of Statistics* **60**, 274–292.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Mathematics* **42**, 43–47.

Scaccia, L. and Martin, R. J. (2005). Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning & Inference* **131**, 19–39.

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*. New York: Chapman & Hall/CRC.

Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society, Series B* **58**, 509–523.

Sherman, M. and Carlstein, E. (1994). Nonparametric estimation of the moments of a general statistic computed from spatial data. *Journal of the American Statistical Association* **89**, 496–500.

Stein, M. L. (1986). A modification of minimum norm quadratic estimation of a generalized covariance function for use with large data sets. *Mathematical Geology* **18**, 625–633.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association* **100**, 310–321.

Taylor, G. I. (1938). The spectrum of turbulance. *Proceedings of the Royal Society of London Ser. A* **164**, 476–490.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001). Spatiotemporal hierarchical bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96**, 382–397.

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92**, 921–936.

APPENDIX A

DERIVATION OF TRANSFORMATION BIAS IN CHAPTER II, SECTION 2.3.3

Suppose X_{s_0} is observed, when $W(s_0) > \widehat{\beta}_2$, plugging $\xi(W) = 1$ and $k = 3$ into model (2.1), we get

$$X(s_0)^{1/3} = \beta_0 + W(s_0)^{1/3}\beta_1 + \epsilon(s_0),$$

where $\epsilon(s_0) \sim \text{Normal}(0, \sigma^2)$. Thus $\text{E}\{\epsilon(s_0)\} = 0$, $\text{E}\{\epsilon(s_0)^3\} = 0$ and $\text{E}\{\epsilon(s_0)^2\} = \text{var}\{\epsilon(s_0)\}$.

However X_{s_0} is not observed, so we estimate it as in (2.5). Under the assumption that β_0 and β_1 are known parameters,

$$\widehat{X}(s_0)^{1/3} = \beta_0 + W(s_0)^{1/3}\beta_1 + \widehat{\epsilon}(s_0),$$

where $\widehat{\epsilon}(s_0) = \sum_{i=1}^{n} \lambda_i \epsilon(s_i)$, so $\text{E}\{\widehat{\epsilon}(s_0)\} = \sum_{i=1}^{n} \lambda_i \text{E}\{\epsilon(s_i)\} = 0$.

Let $f\{W(s_0), \beta\} = \beta_0 + W(s_0)^{1/3}\beta_1$. $f\{W(s_0)\}$ can be generalized to any linear or nonlinear function of $W(s_0)$ and known coefficients.

The derivation here is analogous to the one shown by Cressie (1993, p. 135) in finding the bias for a lognormal random field. When we transform the predictor back by the cubic power transformation, the mean of the true value should be

$$\begin{aligned}
& \text{E}\{(X(s_0)^{1/3})\}^3 \\
= \; & \text{E}([f\{W(s_0), \beta\} + \epsilon(s_0)]^3) \\
= \; & \text{E}([f\{W(s_0), \beta\}]^3 + 3[f\{W(s_0), \beta\}]^2\epsilon(s_0) + 3f\{W(s_0), \beta\}\{\epsilon(s_0)\}^2 + \{\epsilon(s_0)\}^3) \\
= \; & \text{E}([f\{W(s_0), \beta\}]^3 + 3f\{W(s_0), \beta\}\{\epsilon(s_0)\}^2) \\
= \; & [f\{W(s_0), \beta\}]^3 + 3f\{W(s_0), \beta\}\text{var}\{\epsilon(s_0)\}.
\end{aligned}$$

While the mean of our predictor is

$$\begin{aligned}
& \mathrm{E}\{(\widehat{X}(s_0)^{1/3})\}^3 \\
= \; & \mathrm{E}([f\{W(s_0),\beta\}+\widehat{\epsilon}(s_0)]^3) \\
= \; & \mathrm{E}([f\{W(s_0),\beta\}]^3+3[f\{W(s_0),\beta\}]^2\widehat{\epsilon}(s_0)+3f\{W(s_0),\beta\}\{\widehat{\epsilon}(s_0)\}^2+\{\widehat{\epsilon}(s_0)\}^3) \\
= \; & \mathrm{E}([f\{W(s_0),\beta\}]^3+3f\{W(s_0),\beta\}\{\widehat{\epsilon}(s_0)\}^2) \\
= \; & [f\{W(s_0),\beta\}]^3+3f\{W(s_0),\beta\}\mathrm{var}\{\widehat{\epsilon}(s_0)\}.
\end{aligned}$$

The bias induced by the cubic transformation is

$$\mathrm{E}\{(\widehat{X}(s_0)^{1/3})\}^3-\mathrm{E}\{(X(s_0)^{1/3})\}^3=3f\{W(s_0),\beta\}\mathrm{var}\{\widehat{\epsilon}(s_0)\}-\mathrm{var}\{\epsilon(s_0)\},$$

where

$$\begin{aligned}
\mathrm{var}\{\widehat{\epsilon}(s_0)\}-\mathrm{var}\{\epsilon(s_0)\} &= \mathrm{var}\{\widehat{\epsilon}(s_0)-\epsilon(s_0)\}-2[\mathrm{var}\{\epsilon(s_0)\}-\mathrm{cov}\{\widehat{\epsilon}(s_0),\epsilon(s_0)\}] \\
&= \sigma^2(s_0)-2[\mathrm{var}\{\epsilon(s_0)\}-\mathrm{cov}\{\textstyle\sum_{i=1}^n\lambda_i\epsilon(s_i),\epsilon(s_0)\}] \\
&= \sigma^2(s_0)-2[\mathrm{var}\{\epsilon(s_0)\}-\textstyle\sum_{i=1}^n\lambda_i\mathrm{cov}\{\epsilon(s_i),\epsilon(s_0)\}] \\
&= \sigma^2(s_0)-2\textstyle\sum_{i=1}^n\lambda_i[\mathrm{var}\{\epsilon(s_0)\}-\mathrm{cov}\{\epsilon(s_i),\epsilon(s_0)\}] \\
&= \sigma^2(s_0)-2\textstyle\sum_{i=1}^n\lambda_i\gamma(s_i,s_0) \\
&= \sigma^2(s_0)-2\boldsymbol{\lambda}^{\mathrm{T}}\gamma_0
\end{aligned}$$

APPENDIX B

LEMMAS AND PROOF OF THEOREMS IN CHAPTER III

B.1 Lemmas

Lemma B.1.1. *Consider two closed and connected sets U, V in $\mathbb{R}^d$ such that $|U| = |V| \leq b$ and $d(U,V) \geq r$. Let X and Y be measurable random variables with respect to $\mathfrak{F}(U)$ and $\mathfrak{F}(V)$ such that $|X| < C_1$ and $|Y| < C_2$. Then* $\mathrm{cov}(X,Y) \leq 4C_1C_2\alpha_b(r)$.

Proof. The proof is completely analogous to that of Theorem 17.2.1 in Ibragimov and Linnik (1971, p. 306). Note: if the variables X, Y are complex, then separating the real and imaginary parts, we again arrive at the same expression, with 4 replaced by 16.

□

Lemma B.1.2. *Let $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ be a strictly stationary random field which is observed at lattice points in $D_n \subset \mathbb{Z}^d$. Let π be a parameter of the random field and $\widehat{\pi}$ be a consistent estimator of the form $\widehat{\pi}_n = \frac{1}{|D_n|}\sum_{D_n} f\{Z(\mathbf{x})\}$ satisfying $\mathrm{E}[f\{Z(\mathbf{x})\}] = \pi$. Assume*

$$\sum_{\mathbf{x}\in\mathbb{Z}^d} |\mathrm{cov}[f\{Z(\mathbf{0})\}, f\{Z(\mathbf{x})\}]| < \infty.$$

Then $\sigma^2 = \lim_{n\to\infty} |D_n|\mathrm{var}(\widehat{\pi}_n)$ exists. Further assume conditions (C1), (C2) and (C3), setting $p = 0$, $q = d$, $\mathcal{I}_n = D_n$, $\widehat{C}_n(\mathbf{k}) = \widehat{\pi}_n$ and $C(\mathbf{k}) = \pi$, then $\sqrt{|D_n|}(\widehat{\pi}_n - \pi) \xrightarrow{d} \mathrm{N}\ (0, \sigma^2)$.

Proof. Let $A_n = \sqrt{|D_n|}(\widehat{\pi}_n - \pi)$. Then $\sigma^2 = \lim_{n\to\infty} \mathrm{var}(A_n) = \sum_{\mathbf{x}\in\mathbb{Z}^d} \mathrm{cov}[f\{Z(\mathbf{0})\}, f\{Z(\mathbf{x})\}]$ exists by assumption. We apply a blocking technique (e.g., Ibragimov and Linnik, 1971) in conjunction with the mixing condition to prove the normality of $\widehat{\pi}_n$.

Let $l(n) = n^\alpha$ and let $m(n) = n^\alpha - n^\eta$, for some $2d/(d+\epsilon) < \eta < \alpha < 1$. Divide the original field D_n into nonoverlapping subhypercubes, $D^i_{l(n)} = l^d(n)$, $i = 1, ..., k_n$; within each subhypercube, further obtain $D^i_{m(n)}$ which shares the same center as $D^i_{l(n)}$. Thus $d(D^i_{m(n)}, D^{i'}_{m(n)}) \geq n^\eta$ for $i \neq i'$. Figure 16 shows an example when $d = 3$. Let $\widehat{\pi}^i_{m(n)}$ denote the estimator obtained from $D^i_{m(n)}$. Let $a_n = \sum_{i=1}^{k_n} a^i_n/\sqrt{k_n}$, $a'_n = \sum_{i=1}^{k_n}(a^i_n)'/\sqrt{k_n}$, where $a^i_n = \sqrt{m^d(n)}\{\widehat{\pi}^i_{m(n)} - \pi\}$ and $(a^i_n)'$ have the same marginal distributions as a^i_n but are independent. Let $\phi'_n(t)$ and $\phi_n(t)$ be the characteristic functions of a'_n and a_n respectively. The proof consists of the following three steps:

S1 : $A_n - a_n \xrightarrow{p} 0$;

S2 : $\phi'_n(x) - \phi_n(x) \to 0$;

S3 : $a'_n \xrightarrow{d} N(0, \sigma^2)$.

Proof of S1: Since $\mathrm{E}(A_n - a_n) = 0$, it suffices to show $\mathrm{var}(A_n - a_n) \to 0$. Let $D^{m(n)}$ denote the union of all $D^i_{m(n)}$. Observe that

$$\begin{aligned} a_n &= \tfrac{1}{\sqrt{k_n}} \textstyle\sum_{i=1}^{k_n} a^i_n \\ &= \tfrac{\sqrt{m^d(n)}}{\sqrt{k_n}|D^i_{m(n)}|} \textstyle\sum_{i=1}^{k_n} \sum_{\mathbf{x}\in D^i_{m(n)}} [f\{Z(\mathbf{x})\} - \pi] \\ &= \tfrac{\sqrt{k_n}\sqrt{m^d(n)}}{|D^{m(n)}|} \textstyle\sum_{\mathbf{x}\in D^{m(n)}} [f\{Z(\mathbf{x})\} - \pi] \\ &= \sqrt{k_n}\sqrt{m^d(n)}(\widehat{\pi}_{D^{m(n)}} - \pi) \\ &= \sqrt{|D^{m(n)}|}(\widehat{\pi}_{D^{m(n)}} - \pi). \end{aligned}$$

Thus $\mathrm{var}(a_n) \to \sigma^2$ since $\frac{|D_n|}{|D^{m(n)}|} \to 1$ (demonstrated in Lemma B.1.3).

$$\mathrm{cov}(A_n, a_n) = \frac{\sqrt{|D_n||D^{m(n)}|}}{|D_n||D^{m(n)}|} \sum_{\mathbf{x}_1\in D_n} \sum_{\mathbf{x}_2\in D^{m(n)}} \mathrm{cov}[f\{Z(\mathbf{x}_1)\}, f\{Z(\mathbf{x}_2)\}].$$

Again by $\frac{|D_n|}{|D^{m(n)}|} \to 1$, we get $\mathrm{cov}(A_n, a_n) \to \sigma^2$ and $\mathrm{var}(A_n - a_n) \to 0$.

Proof of S2: We employ telescope arguments here. Let ι denote the imaginary number. We define $U_i = \exp(\iota x \frac{a^i_n}{\sqrt{k_n}})$, $X_j = \prod_{i=1}^{j} U_i$, and $Y_j = U_{j+1}$. Completely analogously

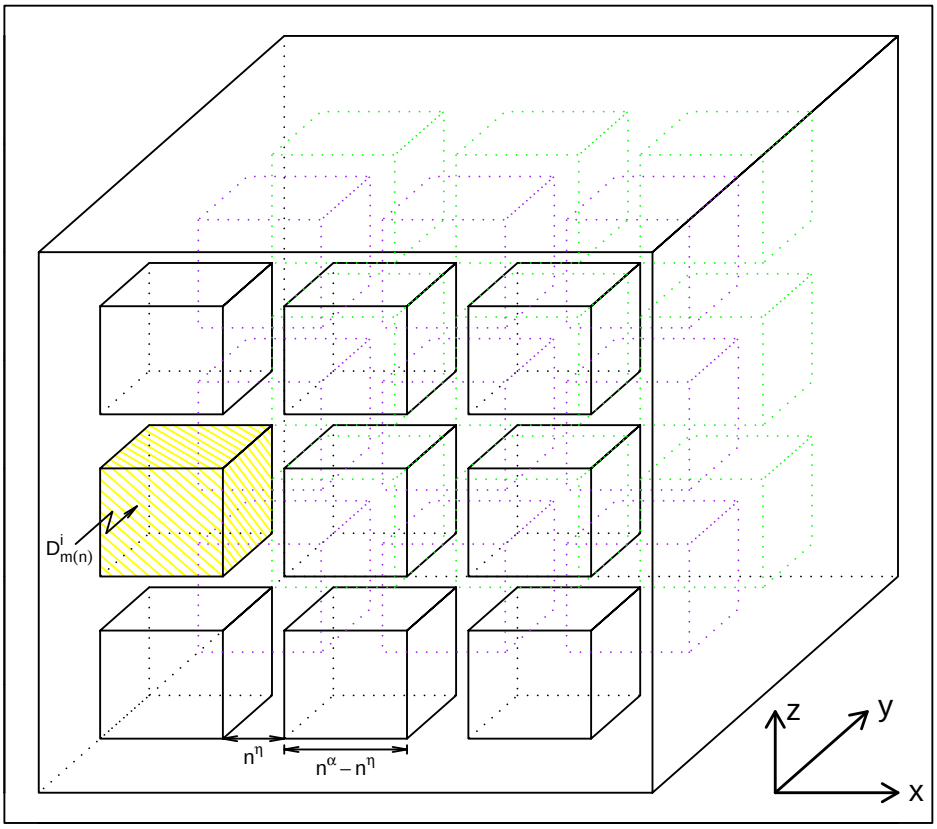

Figure 16: Partition of the random field for Lemma B.1.2

to Guan (2003), we have

$$\mathrm{cov}(X_j, Y_j) \leq 16jm^d(n)n^{-\epsilon\eta} = 16j(n^\alpha - n^\eta)^d n^{-\epsilon\eta} \leq 16jn^{d\alpha-\epsilon\eta}$$

by (C1). Thus

$$|\phi_n'(x) - \phi_n(x)| \leq \sum_{j=1}^{k_n-1} 16jn^{d\alpha-\epsilon\eta} = \mathrm{O}(n^{2d-d\alpha-\epsilon\eta})$$

The last equality in the above expression follows from $\mathrm{O}(k_n)$=$\mathrm{O}(\frac{n^d}{n^{d\alpha}})$=$\mathrm{O}(n^{d(1-\alpha)})$, and $\sum_{j=1}^{k_n-1} 16j = \mathrm{O}(k_n^2)$. Since $2d/(d+\epsilon) < \eta < \alpha < 1$, $2d - d\alpha - \epsilon\eta < 2d - d\eta - \epsilon\eta < 0$. Then $|\phi_n(x) - \phi'_n(x)| \to 0$.

Proof of S3: Observe that $\mathrm{E}(|(a_n^i)'|^{2+\delta}) < C_\delta$ for some constant C_δ. Since $(a_n^i)'$ are i.i.d.,

$$\mathrm{var}\{\sum_{i=1}^{k_n}(a_n^i)'\} = k_n \mathrm{var}\{(a_n^i)'\}.$$

Defining $\sigma_n^2 = \mathrm{var}\{(a_n^i)'\}$, we have $\sigma_n^2 \to \sigma^2$ from the proof of S1. Thus

$$\lim_{n\to\infty} \sum_{i=1}^{k_n} \frac{\mathrm{E}(|(a_n^i)'|^{2+\delta})}{\sqrt{[\mathrm{var}\{\sum_{i=1}^{k_n}(a_n^i)'\}]^{2+\delta}}} \leq \lim_{n\to\infty} C_\delta \frac{k_n}{(k_n\sigma_n^2)^{(2+\delta)/2}} = 0.$$

Thus applying Lyapounov's Theorem, we have

$$\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n}(a_n^i)' \xrightarrow{d} N(0, \sigma^2).$$

□

Lemma B.1.3. *Assuming that D_n satisfies (C2) setting $\mathcal{I}_n = D_n$ and $q = d$, we have $|D^{m(n)}|/|D_n| \to 1$ as $n \to \infty$, where $D^{m(n)}$ is as defined in the proof of Lemma B.1.2.*

Proof. Introduce the following notations, $D^{l(n)}$: the union of $D^i_{l(n)}$ defined in the proof of Lemma B.1.2; $D_{n/l(n)}$: the field in D_n but not in $D^{l(n)}$; $D_{l(n)/m(n)}$: the field in $D^{l(n)}$ but not in $D^{m(n)}$; $k'(n)$: the minimal number of extra $l^d(n)$ subhypercubes needed to cover the whole D_n; $D_{n'}$: the union of all extra $l^d(n)$ subhypercubes needed to cover the whole D_n. Since

$$|D^{l(n)}| = k_n|D^i_{l(n)}| = k_n l^d(n) = k_n n^{d\alpha},$$

$$|D^{m(n)}| = k_n|D^i_{m(n)}| = k_n m^d(n) = k_n(n^\alpha - n^\eta)^d = k_n n^{d\alpha} + \mathrm{o}(k_n n^{d\alpha})$$

noting that $\eta < \alpha$, we have $|D_{l(n)/m(n)}| = |D^{l(n)}| - |D^{m(n)}|$ is $\mathrm{o}(|D^{l(n)}|)$.

We now show that $|D_{n/l(n)}|$ is $o(n^d)$. Note that $D_{n/l(n)} \subset D_{n'}$, i.e., $|D_{n/l(n)}| < |D_{n'}|$. Thus $|D_{n'}|$ is $o(n^d)$ is sufficient for $|D_{n/l(n)}|$ to be $o(n^d)$.

To show this, we split the boundary of D_n into hypercubes of volume $l^{d-1}(n)$. Here we use $L_{n,i}$ to denote the i-th hypercube and k''_n denote the number of all the hypercubes available. k''_n is $O(\frac{n^{d-1}}{l^{d-1}(n)})$ due to (C2). For each $L_{n,i}$, we may form a $l^d(n)$ hypercube (denoted by $HC_{l(n),i}$) which fully contains $L_{n,i}$; in addition to that, we form a $\{3l(n)\}^d$ hypercube (denoted as $HC_{3l(n),i}$) which has the same center as $HC_{l(n),i}$.

Any $l^d(n)$ hypercube that intersects with $L_{n,i}$ (and thus with $HC_{l(n),i}$) will be fully contained in $HC_{3l(n),i}$. Since every subhypercube in $D_{n'}$ intersects with the boundary of D_n and thus with at least one of these $L_{n,i}$s, it will be fully contained in one of the $HC_{3l(n),i}$s. These $l^d(n)$ hypercubes in $D_{n'}$ do not intersect with each other except at the boundary. The maximum number of such hypercubes that intersect with $L_{n,i}$ can not be larger than 3^d. Since the size of k''_n is $O(\frac{n^{d-1}}{l^{d-1}(n)})$, we conclude that k'_n is no larger than $O(\frac{n^{d-1}}{l^{d-1}(n)})$ due to $k'_n < 3^d k''_n$. Thus $|D_{n/l(n)}|$ is no larger than $O(\frac{n^{d-1}}{l^{d-1}(n)} l^d(n))$=$O(n^{d-1}l(n))$=$o(n^d)$. Then

$$\begin{aligned} |D_n| - |D^{m(n)}| &= \{|D_n| - |D^{l(n)}|\} + \{|D^{l(n)}| - |D^{m(n)}|\} \\ &= |D_{n/l(n)}| + |D_{l(n)/m(n)}| \\ &= o(|D_n|) + o(|D_n|) \\ &= o(|D_n|). \end{aligned}$$

Thus $|D^{m(n)}|/|D_n| \to 1$ as $n \to \infty$. □

Lemma B.1.4. *Let $\{Z(\mathbf{s},\mathbf{t}), \mathbf{s} \in \mathbb{R}^p, \mathbf{t} \in \mathbb{R}^q\}$ be a strictly stationary random field which is observed in $D_n = S \times T_n$, where $S \subset \mathbb{R}^p$, $T_n \subset \mathbb{Z}^q$ and $p+q=d$. Let π denote a parameter of the random field and let $\widehat{\pi}$ be a consistent estimator of the form*

$\widehat{\pi}_n = \frac{1}{|S(\pi)|\times|T_n|}\sum_{D_n} f\{Z(\mathbf{s},\mathbf{t})\}$ *satisfying* $\mathrm{E}[f\{Z(\mathbf{s},\mathbf{t})\}] = \pi$, *where* $S(\pi)$ *denotes the set of replicates generated from* S *in estimating* π. *Assume*

$$\sum_{\mathbf{t}\in\mathbb{Z}^q} |\mathrm{cov}[f\{Z(\mathbf{s}_1,\mathbf{0})\}, f\{Z(\mathbf{s}_2,\mathbf{t})\}]| < \infty \text{ for all } \mathbf{s}_1,\mathbf{s}_2 \in S,$$

then $\sigma^2 = \lim_{n\to\infty}|T_n|\mathrm{var}(\widehat{\pi}_n)$ *exists. Further assume (C1), (C2) and (C3), setting* $\mathcal{F} = S$, $\mathcal{I}_n = T_n$, $\widehat{C}_n(\mathbf{k}) = \widehat{\pi}_n$ *and* $C(\mathbf{k}) = \pi$, *then* $\sqrt{|T_n|}(\widehat{\pi}_n - \pi) \xrightarrow{d} \mathrm{N}\ (0,\sigma^2)$.

Proof. Let $A_n = \sqrt{|T_n|}(\widehat{\pi}_n - \pi)$. Then $\sigma^2 = \lim_{n\to\infty}\mathrm{var}(A_n) = \frac{1}{|S(\pi)|^2}\sum_{S(\pi)}\sum_{S(\pi)}\sum_{\mathbf{t}\in\mathbb{Z}^q}$ $\mathrm{cov}[f\{Z(\mathbf{x}_1,\mathbf{0})\}, f\{Z(\mathbf{x}_2,\mathbf{t})\}]$ exists by assumption. We again apply the blocking technique and telescope arguments. Here we divide the whole field into blocks along only T_n, that is, there is no division in space S.

Let $l(n) = n^\alpha$ and let $m(n) = n^\alpha - n^\eta$, for some $2q/(q+\epsilon) < \eta < \alpha < 1$. Divide the original field D_n into nonoverlapping subhypercylinders, $D^i_{l(n)} = S\times T^i_{l(n)}$, $i = 1,...,k_n$, where $T^i_{l(n)} = l^q(n)$; within each subhypercylinder, further obtain $D^i_{m(n)} = S\times T^i_{m(n)}$, which shares the same center as $D^i_{l(n)}$. Thus $d(D^i_{m(n)}, D^{i'}_{m(n)}) \geq n^\eta$ for $i \neq i'$. Figure 17 shows an example when $p = 2, q = 1$. Define $\widehat{\pi}^i_{m(n)}$, a_n, a'_n, $\phi_n(t)$, $\phi'_n(t)$ and the three steps S1, S2 and S3 in the same way as in the proof of Lemma B.1.2, replacing d with q throughout the proof.

For S1, let $D^{m(n)}$ denote the union of all $D^i_{m(n)}$, and $T^{m(n)}$ denote the union of all $T^i_{m(n)}$. Specifically, $|T^{m(n)}| = k_n m^q(n)$ and $|D^{m(n)}| = |S||T^{m(n)}|$. By replacing the new definitions of $|D^i_{m(n)}|$, $|T^{m(n)}|$ and $|D^{m(n)}|$ in S1 of Lemma B.1.2 and observing $\frac{|T_n|}{|T^{m(n)}|} \to 1$ and $\frac{|D_n|}{|D^{m(n)}|} \to 1$, we obtain $a_n = \sqrt{|T^{m(n)}|}(\widehat{\pi}_{D^{m(n)}} - \pi)$, and consequently, $\mathrm{cov}(A_n, a_n) \to \sigma^2$ and $\mathrm{var}(A_n - a_n) \to 0$.

To prove S2, we follow the discussion in S2 of Lemma B.1.2, but replace $jm^d(n)$ with $j|S|m^q(n)$. We have $\mathrm{cov}(X_j, Y_j) \leq 16j|S|n^{q\alpha}n^{-\epsilon\eta}$ and $|\phi'_n(x) - \phi_n(x)| \to 0$ as $n \to \infty$.

The proof of S3 follows directly as in S3 in the proof of Lemma B.1.2.

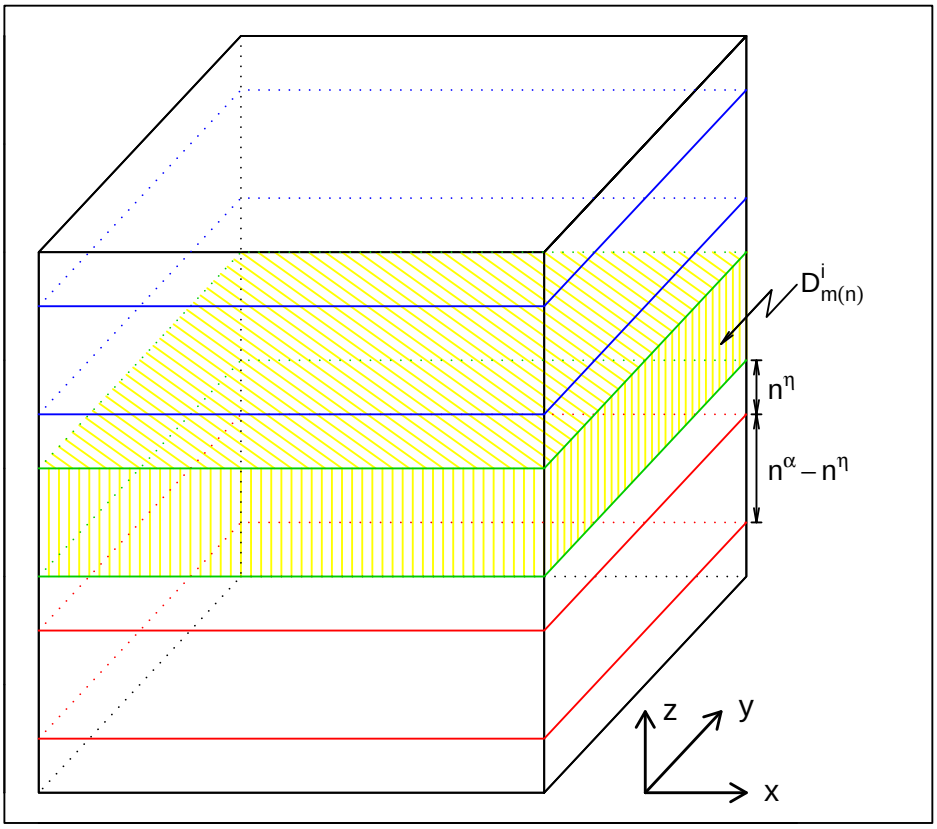

Figure 17: Partition of the random field for Lemma B.1.4

□

Lemma B.1.5. *Let* $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ *be a strictly stationary random field which is observed in* $D_n = S_n \times T_n$*, where* $S_n \subset \mathbb{R}^p$*,* $T_n \subset \mathbb{Z}^q$ *and* $p + q = d$*. Assume the locations of observations in* T_n *are regularly spaced and the locations in* S_n *are generated by a homogeneous Poisson process. Let* $|S_n|$ *denote the volume of* S_n *and*

$|T_n|$ denote the cardinality of T_n. Let

$$\widehat{C}_n(\mathbf{k}) = \frac{1}{\nu^2|T_n||S_n|}\sum_{\mathbf{x}_q\in T_n}\iint_{S_n} w_n(\mathbf{k}_p-\mathbf{x}_{1p}+\mathbf{x}_{2p})Z(\mathbf{x}_{1p},\mathbf{x}_q)Z(\mathbf{x}_{2p},\mathbf{x}_q+\mathbf{k}_q)N^{(2)}(d\mathbf{x}_{1p},d\mathbf{x}_{2p}),$$

where $\mathbf{k}=(\mathbf{k}_p^{\mathrm{T}},\mathbf{k}_q^{\mathrm{T}})^{\mathrm{T}}$, and $\mathbf{x}=(\mathbf{x}_p^{\mathrm{T}},\mathbf{x}_q^{\mathrm{T}})^{\mathrm{T}}$. Assume

$$\sup_{\mathbf{k}_q}\int_{\mathbb{R}^p} C(\mathbf{k}_p,\mathbf{k}_q)d\mathbf{k}_p<\infty,$$

$$\sup_{\mathbf{x}_p,\mathbf{x}_q,\mathbf{x}_{2q}}\sum_{\mathbf{x}_{1q}\in\mathbb{Z}^q}\mathrm{E}\{Z(\mathbf{0},\mathbf{0})Z(\mathbf{x}_p,\mathbf{x}_{1q})Z(\mathbf{0},\mathbf{x}_{2q})Z(\mathbf{x}_p,\mathbf{x}_{2q}+\mathbf{x}_q)\}<\infty,$$

$$\sup_{\mathbf{x}_{1p},\mathbf{x}_{3p}}\sup_{\mathbf{x}_{1q},\mathbf{x}_{2q},\mathbf{x}_{3q}}\int_{\mathbf{x}_{2p}\in\mathbb{R}^p}|Q\{(\mathbf{x}_{1p},\mathbf{x}_{1q}),(\mathbf{x}_{2p},\mathbf{x}_{2q}),(\mathbf{x}_{3p},\mathbf{x}_{3q})\}|d\mathbf{x}_{2p}<\infty,$$

where $w_n(\mathbf{x})=\frac{1}{\lambda_n^p}w(\frac{\mathbf{x}}{\lambda_n})$, and λ_n is a sequence of positive constants satisfying $\lambda_n\to 0$ and $\lambda_n^p|S_n|\to\infty$. Q is defined as in (3.7). Then $\mathrm{E}\{\widehat{C}(\mathbf{k})\}\to C(\mathbf{k})$ and $\Sigma=\lim_{n\to\infty}|S_n|\times|T_n|\lambda_n^p\mathrm{cov}(\widehat{\mathbf{G}}_n,\widehat{\mathbf{G}}_n)$ exists, the (i,j)-th element of which is

$$\tfrac{1}{\nu^2}\textstyle\int w^2(\mathbf{y})d\mathbf{y}\sum_{\mathbf{x}_q\in\mathbb{Z}^q}\mathrm{E}[Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_{ip},\mathbf{k}_{iq})\{Z(\mathbf{0},\mathbf{x}_q)Z(\mathbf{k}_{ip},\mathbf{x}_q+\mathbf{k}_{jq})I(\mathbf{k}_{ip}=\mathbf{k}_{jp})$$
$$+Z(\mathbf{k}_{ip},\mathbf{x}_q)Z(\mathbf{0},\mathbf{x}_q+\mathbf{k}_{jq})I(\mathbf{k}_{ip}=-\mathbf{k}_{jp})\}],$$

where $I(\mathbf{k}_{ip}=\pm\mathbf{k}_{jp})=1$ if $\mathbf{k}_{ip}=\pm\mathbf{k}_{jp}$ and 0 otherwise. If we further assume that Σ is positive definite and conditions (C1), (C2) and (C3′) hold setting $p=0$, $q=d$, $q_1=q$, $\mathcal{I}_n=D_n$ and replacing $C(\mathbf{k})$ with $E\{C_n(\mathbf{k})\}$, then

$$\sqrt{|T_n||S_n|\lambda_n^q}\{\widehat{\mathbf{G}}_n-\mathrm{E}(\widehat{\mathbf{G}}_n)\}\xrightarrow{d}\mathrm{N}_m(\mathbf{0},\Sigma).$$

Proof.

$$\begin{aligned}
&\mathrm{E}\{\widehat{C}_n(\mathbf{k})\}\\
=\;&\tfrac{1}{\nu^2|T_n||S_n|}\sum_{\mathbf{x}_q\in T_n}\mathrm{E}\{\iint_{S_n} w_n(\mathbf{k}_p+\mathbf{x}_{1p}-\mathbf{x}_{2p})Z(\mathbf{x}_{1p},\mathbf{x}_q)Z(\mathbf{x}_{2p},\mathbf{x}_q+\mathbf{k}_q)N^{(2)}(d\mathbf{x}_{1p},d\mathbf{x}_{2p})\}\\
=\;&\tfrac{1}{|T_n|}\textstyle\sum_{\mathbf{x}_q\in T_n}\tfrac{1}{|S_n|}\iint_{S_n} w_n(\mathbf{k}_p+\mathbf{x}_{1p}-\mathbf{x}_{2p})C(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{k}_q)d\mathbf{x}_{1p}d\mathbf{x}_{2p}\\
=\;&\tfrac{1}{|T_n|}\textstyle\sum_{\mathbf{x}_q\in T_n}\tfrac{1}{|S_n|}\iint_{S_n}\tfrac{1}{\lambda_n^p}w(\tfrac{\mathbf{k}_p+\mathbf{x}_{1p}-\mathbf{x}_{2p}}{\lambda_n})C(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{k}_q)d\mathbf{x}_{1p}d\mathbf{x}_{2p}\\
=\;&\tfrac{1}{|T_n|}\textstyle\sum_{\mathbf{x}_q\in T_n}\int_{S_n-S_n} w(\mathbf{y})C(\mathbf{k}_p-\lambda_n\mathbf{y},\mathbf{k}_q)\tfrac{|S_n\cap(S_n-\mathbf{k}_p+\lambda_n\mathbf{y})|}{|S_n|}d\mathbf{y},
\end{aligned}$$

where $S_n - S_n$ denotes all pairwise differences between the two sets. Since

$$\begin{aligned} & \textstyle\int_{S_n - S_n} w(\mathbf{y}) C(\mathbf{k}_p - \lambda_n \mathbf{y}, \mathbf{k}_q) \frac{|S_n \cap (S_n - \mathbf{k}_p + \lambda_n \mathbf{y})|}{|S_n|} d\mathbf{y} \\ \longrightarrow \quad & C(\mathbf{k}_p, \mathbf{k}_q) \textstyle\int_{\mathbb{R}^q} w(\mathbf{y}) d\mathbf{y} = C(\mathbf{k}_p, \mathbf{k}_q), \end{aligned}$$

we have $\mathrm{E}\{\widehat{C}_n(\mathbf{k})\} \longrightarrow C(\mathbf{k})$.

The derivation for the variance is analogous to Karr (1986). Specifically, consider two spatial locations $\mathbf{k} = (\mathbf{k}_p^{\mathrm{T}}, \mathbf{k}_q^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{k}' = ({\mathbf{k}'_p}^{\mathrm{T}}, {\mathbf{k}'_q}^{\mathrm{T}})^{\mathrm{T}}$:

$$\begin{aligned} & \mathrm{E}\{\widehat{C}_n(\mathbf{k}) \widehat{C}_n(\mathbf{k}')\} \\ = \; & \tfrac{1}{\nu^4 |T_n|^2} \textstyle\sum_{\mathbf{x}_q \in T_n} \sum_{\mathbf{x}'_q \in T_n} \frac{1}{|S_n|^2} \iiiint_{S_n} w_n(\mathbf{k}_p + \mathbf{x}_{1p} - \mathbf{x}_{2p}) w_n(\mathbf{k}'_p + \mathbf{x}'_{1p} - \mathbf{x}'_{2p}) \\ & \times \mathrm{E}\{Z(\mathbf{x}_{1p}, \mathbf{x}_q) Z(\mathbf{x}_{2p}, \mathbf{x}_q + \mathbf{k}_q) Z(\mathbf{x}'_{1p}, \mathbf{x}'_q) Z(\mathbf{x}'_{2p}, \mathbf{x}'_q + \mathbf{k}'_q)\} \\ & \times \mathrm{E}\{N^{(2)}(d\mathbf{x}_{1p}, d\mathbf{x}_{2p}) N^{(2)}(d\mathbf{x}'_{1p}, d\mathbf{x}'_{2p})\}. \end{aligned} \tag{B.1}$$

Observe that

$$\begin{aligned} & \mathrm{E}\{Z(\mathbf{x}_{1p}, \mathbf{x}_q) Z(\mathbf{x}_{2p}, \mathbf{x}_q + \mathbf{k}_q) Z(\mathbf{x}'_{1p}, \mathbf{x}'_q) Z(\mathbf{x}'_{2p}, \mathbf{x}'_q + \mathbf{k}'_q)\} \\ = \; & Q\{(\mathbf{x}_{2p} - \mathbf{x}_{1p}, \mathbf{k}_q), (\mathbf{x}'_{1p} - \mathbf{x}_{1p}, \mathbf{x}'_q - \mathbf{x}_q), (\mathbf{x}'_{2p} - \mathbf{x}_{1p}, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q)\} \\ & + C(\mathbf{x}_{2p} - \mathbf{x}_{1p}, \mathbf{k}_q) C(\mathbf{x}'_{2p} - \mathbf{x}'_{1p}, \mathbf{k}'_q) \\ & + C(\mathbf{x}'_{1p} - \mathbf{x}_{1p}, \mathbf{x}'_q - \mathbf{x}_q) C(\mathbf{x}'_{2p} - \mathbf{x}_{2p}, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q - \mathbf{k}_q) \\ & + C(\mathbf{x}'_{2p} - \mathbf{x}_{1p}, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q) C(\mathbf{x}_{2p} - \mathbf{x}'_{1p}, \mathbf{x}_q - \mathbf{x}'_q + \mathbf{k}_q). \end{aligned} \tag{B.2}$$

In addition,

$$\begin{aligned} & \mathrm{E}\{N^{(2)}(d\mathbf{x}_{1p}, d\mathbf{x}_{2p}) N^{(2)}(d\mathbf{x}'_{1p}, d\mathbf{x}'_{2p})\} \\ = \; & \nu^4 d\mathbf{x}_{1p} d\mathbf{x}_{2p} d\mathbf{x}'_{1p} d\mathbf{x}'_{2p} + \nu^3 d\mathbf{x}_{1p} d\mathbf{x}_{2p} \epsilon_{\mathbf{x}_{1p}}(d\mathbf{x}'_{1p}) d\mathbf{x}'_{2p} \\ & + \nu^3 d\mathbf{x}_{1p} d\mathbf{x}_{2p} d\mathbf{x}'_{1p} \epsilon_{\mathbf{x}_{1p}}(d\mathbf{x}'_{2p}) \\ & + \nu^3 d\mathbf{x}_{1p} d\mathbf{x}_{2p} \epsilon_{\mathbf{x}_{2p}}(d\mathbf{x}'_{1p}) d\mathbf{x}'_{2p} + \nu^3 d\mathbf{x}_{1p} d\mathbf{x}_{2p} d\mathbf{x}'_{1p} \epsilon_{\mathbf{x}_{2p}}(d\mathbf{x}'_{2p}) \\ & + \nu^2 d\mathbf{x}_{1p} d\mathbf{x}_{2p} \epsilon_{\mathbf{x}_{1p}}(d\mathbf{x}'_{2p}) \epsilon_{\mathbf{x}_{1p}}(d\mathbf{x}'_{2p}) + \nu^2 d\mathbf{x}_{1p} d\mathbf{x}_{2p} \epsilon_{\mathbf{x}_{2p}}(d\mathbf{x}'_{1p}) \epsilon_{\mathbf{x}_{2p}}(d\mathbf{x}'_{1p}), \end{aligned} \tag{B.3}$$

where $\epsilon_{\mathbf{x}}(\cdot)$ denotes point measure.

Substituting these two expansions into the main formula produces a lengthy expression that we do not reproduce in its entirety. We only show some representative terms rather than the total 28 terms. For simplicity, we first look at only the integral part coditional on $\mathbf{x}^q$ and $\mathbf{x}'_q$. Expanding (B.1) first by (B.2) and then by (B.3), the first term is

$$
\begin{aligned}
& \tfrac{1}{\nu^4|S_n|^2} \iiiint\limits_{S_n} w_n(\mathbf{k}_p + \mathbf{x}_{1p} - \mathbf{x}_{2p}) w_n(\mathbf{k}'_p + \mathbf{x}'_{1p} - \mathbf{x}'_{2p}) Q\{(\mathbf{x}_{2p} - \mathbf{x}_{1p}, \mathbf{k}_q), \\
& (\mathbf{x}'_{1p} - \mathbf{x}_{1p}, \mathbf{x}'_q - \mathbf{x}_q), (\mathbf{x}'_{2p} - \mathbf{x}_{1p}, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q)\} \nu^4 d\mathbf{x}_{1p} d\mathbf{x}_{2p} d\mathbf{x}'_{1p} d\mathbf{x}'_{2p} \\
= & \iiint\limits_{S_n - S_n} w_n(\mathbf{k}_p - \mathbf{v}_1) w_n(\mathbf{k}'_p + \mathbf{v}_2 - \mathbf{v}_3) Q\{(\mathbf{v}_1, \mathbf{k}_q), (\mathbf{v}_2, \mathbf{x}'_q - \mathbf{x}_q), (\mathbf{v}_3, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q)\} \\
& \times \tfrac{|S_n \cap (S_n - \mathbf{v}_1) \cap (S_n - \mathbf{v}_2) \cap (S_n - \mathbf{v}_3)|}{|S_n|^2} d\mathbf{v}_1 d\mathbf{v}_2 d\mathbf{v}_3 \\
\leq & \iiint\limits_{S_n - S_n} w_n(\mathbf{k}_p - \mathbf{v}_1) w_n(\mathbf{k}'_p + \mathbf{v}_2 - \mathbf{v}_3) | Q\{(\mathbf{v}_1, \mathbf{k}_q), (\mathbf{v}_2, \mathbf{x}'_q - \mathbf{x}_q), (\mathbf{v}_3, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q)\} \\
& | \tfrac{1}{|S_n|} d\mathbf{v}_1 d\mathbf{v}_2 d\mathbf{v}_3 \\
\leq & \iiint\limits_{S_n - S_n} w_n(\mathbf{k}_p - \mathbf{v}_1) w_n(\mathbf{k}'_p - \mathbf{v}_4) | Q\{(\mathbf{v}_1, \mathbf{k}_q), (\mathbf{v}_2, \mathbf{x}'_q - \mathbf{x}_q), (\mathbf{v}_2 + \mathbf{v}_4, \mathbf{x}'_q - \mathbf{x}_q + \mathbf{k}'_q)\} \\
& | \tfrac{1}{|S_n|} d\mathbf{v}_1 d\mathbf{v}_2 d\mathbf{v}_4 \\
\leq & C_1 \iint\limits_{\mathbb{R}^p} w_n(\mathbf{k}_p - \mathbf{v}_1) w_n(\mathbf{k}'_p - \mathbf{v}_4) \tfrac{1}{|S_n|} d\mathbf{v}_1 d\mathbf{v}_4 \\
= & \mathrm{O}(\tfrac{1}{|S_n|}).
\end{aligned}
$$

The second term is

$$
\begin{aligned}
& \tfrac{1}{\nu^4|S_n|^2} \iiiint\limits_{S_n} w_n(\mathbf{k}_p + \mathbf{x}_{1p} - \mathbf{x}_{2p}) w_n(\mathbf{k}'_p + \mathbf{x}'_{1p} - \mathbf{x}'_{2p}) \\
& \times C(\mathbf{x}_{2p} - \mathbf{x}_{1p}, \mathbf{k}_q) C(\mathbf{x}'_{2p} - \mathbf{x}'_{1p}, \mathbf{k}'_q) \nu^4 d\mathbf{x}_{1p} d\mathbf{x}_{2p} d\mathbf{x}'_{1p} d\mathbf{x}'_{2p} \\
= & \tfrac{1}{|S_n|^2} \iint\limits_{S_n - S_n} w_n(\mathbf{k}_p + \mathbf{x}_{1p} - \mathbf{x}_{2p}) C(\mathbf{x}_{2p} - \mathbf{x}_{1p}, \mathbf{k}_q) d\mathbf{x}_{1p} d\mathbf{x}_{2p} \\
& \times \iint\limits_{S_n - S_n} w_n(\mathbf{k}'_p + \mathbf{x}'_{1p} - \mathbf{x}'_{2p}) C(\mathbf{x}'_{2p} - \mathbf{x}'_{1p}, \mathbf{k}'_q) d\mathbf{x}'_{1p} d\mathbf{x}'_{2p} \\
= & \textstyle\int_{S_n - S_n} w_n(\mathbf{k}_p - \mathbf{v}_1) C(\mathbf{v}_1, \mathbf{k}_q) \tfrac{|S_n \cap (S_n - \mathbf{v}_1)|}{|S_n|} d\mathbf{v}_1 \int_{S_n - S_n} w_n(\mathbf{k}'_p - \mathbf{v}_2) C(\mathbf{v}_2, \mathbf{k}'_q) \\
& \times \tfrac{|S_n \cap (S_n - \mathbf{v}_2)|}{|S_n|} d\mathbf{v}_2 \\
\to & C(\mathbf{k}) \textstyle\int_{\mathbb{R}^p} w(\mathbf{y}) d\mathbf{y} \times C(\mathbf{k}') \int_{\mathbb{R}^p} w(\mathbf{y}) d\mathbf{y} \\
= & C(\mathbf{k})|_{\mathbf{x}_q} \times C(\mathbf{k}')|_{\mathbf{x}'_q}.
\end{aligned}
$$

where $C(\mathbf{k})|_{\mathbf{x}_q}$ denotes $C(\mathbf{k})$ conditional on $\mathbf{x}_q$. Notice that

$\frac{1}{|T_n|^2}\sum_{\mathbf{x}_q\in T_n}\sum_{\mathbf{x}'_q\in T_n} C(\mathbf{k})|_{\mathbf{x}_q}\times C(\mathbf{k}')|_{\mathbf{x}'_q} \longrightarrow C(\mathbf{k})C(\mathbf{k}')$.

Most of the terms are smaller or the same as $\mathrm{O}(\frac{1}{|S_n|})$. The terms that are larger than these are the dominant contributions among the 28 terms, which includes the 21-*st* to 24-*th* terms:

$$
\begin{aligned}
&\tfrac{1}{\nu^4|S_n|^2}\iint\limits_{S_n} w_n(\mathbf{k}_p+\mathbf{x}_{1p}-\mathbf{x}_{2p})w_n(\mathbf{k}'_p+\mathbf{x}_{1p}-\mathbf{x}_{2p})\\
&\times[Q\{(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{k}_q),(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q),(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)\}\\
&+C(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{k}_q)C(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{k}'_q)+C(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q)C(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q-\mathbf{k}_q)\\
&+C(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)C(\mathbf{x}_{2p}-\mathbf{x}_{1p},\mathbf{x}_q-\mathbf{x}'_q+\mathbf{k}_q)]\nu^2 d\mathbf{x}_{1p}d\mathbf{x}_{2p}\\
=\;&\tfrac{1}{\nu^2}\textstyle\int_{S_n-S_n} w_n(\mathbf{k}_p-\mathbf{v})w_n(\mathbf{k}'_p-\mathbf{v})\times[Q\{(\mathbf{v},\mathbf{k}_q),(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q),(\mathbf{v},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)\}\\
&+C(\mathbf{v},\mathbf{k}_q)C(\mathbf{v},\mathbf{k}'_q)+C(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q)C(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q-\mathbf{k}_q)\\
&+C(\mathbf{v},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)C(\mathbf{v},\mathbf{x}_q-\mathbf{x}'_q+\mathbf{k}_q)]\tfrac{|S_n\cap(S_n-\mathbf{v})|}{|S_n|^2}d\mathbf{v}\\
=\;&\tfrac{1}{\nu^2\lambda_n^p}\textstyle\int_{S_n-S_n} w(\mathbf{y})w(\mathbf{y}+\frac{\mathbf{k}'_p-\mathbf{k}_p}{\lambda_n})\mathrm{E}\{Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_p-\lambda_n\mathbf{y},\mathbf{k}_q)Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q)\\
&\times Z(\mathbf{k}_p-\lambda_n\mathbf{y},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)\}\tfrac{|S_n\cap(S_n-\mathbf{k}_p-\lambda_n\mathbf{y})|}{|S_n|^2}d\mathbf{y},
\end{aligned}
$$

and the 25-*th* to 28-*th* terms:

$$
\begin{aligned}
&\tfrac{1}{\nu^2\lambda_n^p}\textstyle\int_{S_n-S_n} w(\mathbf{y})w(\mathbf{y}-\frac{\mathbf{k}'_p+\mathbf{k}_p}{\lambda_n})\mathrm{E}\{Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_p-\lambda_n\mathbf{y},\mathbf{k}_q)Z(\mathbf{k}_p-\lambda_n\mathbf{y},\mathbf{x}'_q-\mathbf{x}_q)\\
&\times Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)\}\tfrac{|S_n\cap(S_n-\mathbf{k}_p-\lambda_n\mathbf{y})|}{|S_n|^2}d\mathbf{y}.
\end{aligned}
$$

Combining the 21-*st* to 28-*th* terms, we get

$$
\begin{aligned}
&|S_n|\lambda_n^p\mathrm{cov}\{C(\mathbf{k})|_{\mathbf{x}_q},C(\mathbf{k}')|_{\mathbf{x}'_q}\}\\
\longrightarrow\;&\tfrac{1}{\nu^2}\textstyle\int_{\mathbb{R}^p} w(\mathbf{y})d\mathbf{y}\mathrm{E}[Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_p,\mathbf{k}_q)\{Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q)Z(\mathbf{k}_p,\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=\mathbf{k}'_p)\\
&+Z(\mathbf{k}_p,\mathbf{x}'_q-\mathbf{x}_q)Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=-\mathbf{k}'_p)\}].
\end{aligned}
$$

Let

$$\begin{aligned} X &:= \tfrac{1}{|T_n|^2}\textstyle\sum_{\mathbf{x}_q\in T_n}\sum_{\mathbf{x}'_q\in T_n} \mathrm{E}[Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_p,\mathbf{k}_q)\{Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q)Z(\mathbf{k}_p,\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q) \\ &\quad \times I(\mathbf{k}_p=\mathbf{k}'_p)+Z(\mathbf{k}_p,\mathbf{x}'_q-\mathbf{x}_q)Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=-\mathbf{k}'_p)\}] \\ &= \mathrm{E}[Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_p,\mathbf{k}_q)\tfrac{1}{|T_n|^2}\sum_{T_n}\sum_{T_n}\{Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q)Z(\mathbf{k}_p,\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=\mathbf{k}'_p) \\ &\quad +Z(\mathbf{k}_p,\mathbf{x}'_q-\mathbf{x}_q)Z(\mathbf{0},\mathbf{x}'_q-\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=-\mathbf{k}'_p)\}] \\ &= \mathrm{E}[Z(\mathbf{0},\mathbf{0})Z(\mathbf{k}_p,\mathbf{k}_q)\textstyle\sum_{T_n-T_n}\{Z(\mathbf{0},\mathbf{v})Z(\mathbf{k}_p,\mathbf{v}+\mathbf{k}'_q)I(\mathbf{k}_p=\mathbf{k}'_p) \\ &\quad +Z(\mathbf{k}_p,\mathbf{v})Z(\mathbf{0},\mathbf{v}+\mathbf{k}'_q)I(\mathbf{k}_p=-\mathbf{k}'_p)\}\tfrac{T_n\cap(T_n-\mathbf{v})}{|T_n|^2}]. \end{aligned}$$

Applying Kronecker's Lemma,

$$\begin{aligned} |T_n|X \longrightarrow\ & \textstyle\sum_{\mathbf{x}_q\in\mathbb{Z}^q}\mathrm{E}[Z(\mathbf{0})Z(\mathbf{k})\{Z(\mathbf{0},\mathbf{x}_q)Z(\mathbf{k}_p,\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=\mathbf{k}'_p) \\ & +Z(\mathbf{k}_p,\mathbf{x}_q)Z(\mathbf{0},\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=-\mathbf{k}'_p)\}]. \end{aligned}$$

Thus,

$$\begin{aligned} & |S_n||T_n|\lambda_n^p\mathrm{cov}(C(\mathbf{k})|_{\mathbf{x}_q},C(\mathbf{k}')|_{\mathbf{x}'_q}) \\ \longrightarrow\ & \textstyle\sum_{\mathbf{x}_q\in\mathbb{Z}^q}\mathrm{E}[Z(\mathbf{0})Z(\mathbf{k})\{Z(\mathbf{0},\mathbf{x}_q)Z(\mathbf{k}_p,\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=\mathbf{k}'_p) \\ & +Z(\mathbf{k}_p,\mathbf{x}_q)Z(\mathbf{0},\mathbf{x}_q+\mathbf{k}'_q)I(\mathbf{k}_p=-\mathbf{k}'_p)\}]\tfrac{1}{\nu^2}\textstyle\int_{\mathbb{R}^p}w^2(\mathbf{y})d\mathbf{y}. \end{aligned}$$

There are two different terms for $I(\mathbf{k}_p=\mathbf{k}'_p)$ and $I(\mathbf{k}_p=-\mathbf{k}'_p)$ because $C(\mathbf{k}_p,\mathbf{k}_q)$ does not equal $C(-\mathbf{k}_p,\mathbf{k}_q)$ in general.

The proof of normality again uses the three steps as in Lemma B.1.2. The proof of S1 is analogous to that in Lemma B.1.2 but using a kernel estimator and an additional requirement for α, $\lambda_n^p n^{\alpha p}\to\infty$. We follow Politis, Paparoditis, and Romano (1998) to prove S2. Define X_i and Y_i as in Lemma B.1.2 and

$$\mathrm{E}^N(X_i)=\mathrm{E}(X_i|N),\ \ \mathrm{E}^N(Y_i)=\mathrm{E}(Y_i|N),\ \ \mathrm{cov}^N(X_i,Y_i)=\mathrm{cov}(X_i,Y_i|N).$$

Then

$$\mathrm{cov}(X_i,Y_i)=\mathrm{E}\{\mathrm{cov}^N(X_i,Y_i)\}+\mathrm{cov}\{\mathrm{E}^N(X_i),\mathrm{E}^N(Y_i)\}.$$

Since N is a homogeneous Poisson process and X_i, Y_i are random variables defined on two disjoint random fields, we have

$$\text{cov}\{\text{E}^N(X_i), \text{E}^N(Y_i)\} = 0.$$

For given N, $X_i|N$ is measurable with respect to $\mathfrak{F}(\bigcup_{j=1}^{i} D^j_{m(n)})$ and $Y_i|N$ is measurable with respect to $\mathfrak{F}(D^{i+1}_{m(n)})$. Then we have

$$|\phi'_n(x) - \phi_n(x)| \leq \text{cov}(X_i, Y_i|N).$$

The rest is completely analogous to Lemma B.1.2.

□

Lemma B.1.6. *Let $\{Z(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d\}$ be a strictly stationary random field with mean μ which is observed in $D_n \subset \mathbb{R}^d$ satisfying condition (C2) setting $q = d$ and $\mathcal{I}_n = D_n$. Let $\overline{Z}_n = \frac{1}{|D_n|}\sum_{D_n} Z(\mathbf{x})$. Assume mixing condition (C1) and set $q = d$. If*

$$\sum_{\mathbf{x}\in\mathbb{R}^d} |\text{cov}\{Z(\mathbf{0}), Z(\mathbf{x})\}| < \infty \text{ for all finite } \mathbf{x},$$

and

$$\sup_n \text{E}|\sqrt{|D_n|}(\overline{Z} - \mu)|^{2+\delta} \leq C_\delta, \text{ for some } \delta > 0, \ C_\delta < \infty,$$

then $\sigma^2_{\overline{Z}} = \lim_{n\to\infty} |D_n|\text{var}(\overline{Z}_n)$ exists and $\sqrt{|D_n|}(\overline{Z}_n - \mu) \xrightarrow{d} \text{N}\ (0, \sigma^2_{\overline{Z}})$.

Proof. The proof follows directly from Lemma B.1.2.

□

It is straightforward to prove that the results in Lemma B.1.6 hold for all the situations we have considered in this article given the appropriate mixing condition. We have proved $\widehat{\mathbf{G}}_n = \{\widehat{C}_n(\mathbf{h}, u) : (\mathbf{h}, u) \in \mathbf{\Lambda}\}$ is asymptotically jointly normal, that is, $\sqrt{|D_n|} \times (\widehat{\mathbf{G}}_n - \widehat{\mathbf{G}}) \xrightarrow{d} \text{N}_m(\mathbf{0}, \Sigma)$ under appropriate assumptions. In this formulation of $\widehat{\mathbf{G}}_n$, we have assumed the mean of $Z(\mathbf{s}, t)$, $\mu(Z)$, is known, and without

loss of generality assumed it equals 0. However in most cases, the mean of the random process is unknown and needs to be estimated.

Lemma B.1.7. *Let* $\widehat{\mu}_n(Z) := \overline{Z}_n = \frac{1}{|D_n|}\sum_{\mathbf{x}\in D_n} Z(\mathbf{x})$, *and let* $\widehat{\mathbf{G}}_n^* = \{\widehat{C}_n^*(\mathbf{k}) : \mathbf{k} \in \mathbf{\Lambda}\}$, *where* $\widehat{C}_n^*(\mathbf{k}) = \frac{1}{|D_n(\mathbf{k})|}\sum_{\mathbf{x}\in D_n}\{Z(\mathbf{x}) - \widehat{\mu}_n\}\{Z(\mathbf{x}+\mathbf{k}) - \widehat{\mu}_n\}$. *Then the vector of sample covariances* $\widehat{\mathbf{G}}_n^*$ *has the same asymptotic properties as* $\widehat{\mathbf{G}}_n$, *i.e.,* $\sqrt{|D_n|}(\widehat{\mathbf{G}}_n^* - \mathbf{G}) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, \Sigma)$.

Proof. Assume $\mu \neq 0$. Now we prove $\sqrt{|D_n|}\{\widehat{C}_n(\mathbf{k}) - \widehat{C}_n^*(\mathbf{k})\} = \mathrm{o}_p(1)$ as $n \to \infty$. Following Brockwell and Davis (1991, p. 229)

$$
\begin{aligned}
& \sqrt{|D_n|}\{\widehat{C}_n(\mathbf{k}) - \widehat{C}_n^*(\mathbf{k})\} \\
= \;& \sqrt{|D_n|}\tfrac{1}{|D_n(\mathbf{k})|}\textstyle\sum_{\mathbf{x}\in D_n}(\overline{Z}_n - \mu)\{Z(\mathbf{x}) + Z(\mathbf{x}+\mathbf{k})\} + (\mu^2 - \overline{Z}_n^2) \\
= \;& \sqrt{|D_n|}(\overline{Z}_n - \mu)\tfrac{1}{|D_n(\mathbf{k})|}\textstyle\sum_{\mathbf{x}\in D_n}\{Z(\mathbf{x}) + Z(\mathbf{x}+\mathbf{k}) - \mu - \overline{Z}_n\} \\
= \;& \sqrt{|D_n|}(\overline{Z}_n - \mu)\{\tfrac{1}{|D_n(\mathbf{k})|}\textstyle\sum_{\mathbf{x}\in D_n} Z(\mathbf{x}) - \mu + \tfrac{1}{|D_n(\mathbf{k})|}\sum_{\mathbf{x}\in D_n} Z(\mathbf{x}) - \overline{Z}_n\} \\
= \;& \sqrt{|D_n|}(\overline{Z}_n - \mu)(\overline{Z}_n - \mu).
\end{aligned}
$$

By Lemma B.1.6 we know that $\sqrt{|D_n|}(\overline{Z}_n - \mu) \xrightarrow{d} \mathrm{N}(0, \sigma_{\overline{Z}}^2)$, which implies that $\sqrt{|D_n|}(\overline{Z}_n - \mu)$ is $\mathrm{O}_p(1)$ and $\overline{Z}_n - \mu \xrightarrow{p} 0$.

From these observations we conclude that $\sqrt{|D_n|}(\widehat{\mathbf{G}}_n - \widehat{\mathbf{G}}_n^*) \to \mathrm{o}_p(1)$ as $n \to \infty$. Thus, $\sqrt{|D_n|}(\widehat{\mathbf{G}}_n^* - \mathbf{G}) \xrightarrow{d} \mathrm{N}_m(\mathbf{0}, \Sigma)$.

□

Proof of Lemma 3.4.1

Proof. When Z_t is Gaussian, the fourth order cumulant function $Q = 0$ (Isserlis,

1918).

$$\begin{aligned}
& \operatorname{cov}\{\widehat{C}_{i_1,j_1}(s), \widehat{C}_{i_2,j_2}(u)\} \\
= \; & \mathrm{E}\{\widehat{C}_{i_1,j_1}(s)\widehat{C}_{i_2,j_2}(u)\} - \mathrm{E}\{\widehat{C}_{i_1,j_1}(s)\}\mathrm{E}\{\widehat{C}_{i_2,j_2}(u)\} \\
= \; & \tfrac{1}{|T_n|^2}\textstyle\sum_{t_1}\sum_{t_2}\mathrm{E}(Z_{j_1,t_1}Z_{i_1,t_1+s}Z_{j_2,t_2}Z_{i_2,t_2+u}) - C_{i_1,j_1}(s)C_{i_2,j_2}(u) \\
= \; & \tfrac{1}{|T_n|^2}\textstyle\sum_{t_1}\sum_{t_2}\{C_{i_1,j_1}(s)C_{i_2,j_2}(u) + C_{j_2,j_1}(t_2-t_1)C_{i_2,i_1}(t_2-t_1+u-s) \\
& + C_{i_2,j_1}(t_2-t_1+u)C_{j_2,i_1}(t_2-t_1-s)\} - C_{i_1,j_1}(s)C_{i_2,j_2}(u) + Q.
\end{aligned}$$

So $|T_n|\operatorname{cov}\{\widehat{C}_{i_1,j_1}(s), \widehat{C}_{i_2,j_2}(u)\} \to \sum_{r\in\mathbb{Z}}\{C_{j_2,j_1}(r)C_{i_2,i_1}(r+u-s) + C_{i_2,j_1}(r+u)C_{j_2,i_1}(r-s)\}$ as $n \to \infty$.

Let $T_n(u) = \{t : t \in T_n, t+u \in T_n\}$, we have

$$\begin{aligned}
& \operatorname{cov}\{\widehat{C}_n(\mathbf{h}_i, u_i), \widehat{C}_n(\mathbf{h}_j, u_j)\} \\
= \; & \tfrac{1}{|S(\mathbf{h}_i)||T_n|}\tfrac{1}{|S(\mathbf{h}_j)||T_n|}\textstyle\sum_{\mathbf{s}_1\in S(\mathbf{h}_i)}\sum_{t_1\in T_n(u_i)}\sum_{\mathbf{s}_2\in S(\mathbf{h}_j)}\sum_{t_2\in T_n(u_j)} \\
& \operatorname{cov}\{Z(\mathbf{s}_1,t_1)Z(\mathbf{s}_1+\mathbf{h}_i,t_1+u_i), Z(\mathbf{s}_2,t_2)Z(\mathbf{s}_2+\mathbf{h}_j,t_2+u_j)\} \\
= \; & \tfrac{1}{|S(\mathbf{h}_i)|}\tfrac{1}{|S(\mathbf{h}_j)|}\textstyle\sum_{\mathbf{s}_1\in S(\mathbf{h}_i)}\sum_{\mathbf{s}_2\in S(\mathbf{h}_j)}\operatorname{cov}\{\tfrac{1}{|T_n|}\sum_{t_1\in T_n(u_i)}Z(\mathbf{s}_1,t_1)Z(\mathbf{s}_1+\mathbf{h}_i,t_1+u_i), \\
& \tfrac{1}{|T_n|}\textstyle\sum_{t_2\in T_n(u_j)}Z(\mathbf{s}_2,t_2)Z(\mathbf{s}_2+\mathbf{h}_j,t_2+u_j)\} \\
= \; & \tfrac{1}{|S(\mathbf{h}_i)|}\tfrac{1}{|S(\mathbf{h}_j)|}\textstyle\sum_{\mathbf{s}_1\in S(\mathbf{h}_i)}\sum_{\mathbf{s}_2\in S(\mathbf{h}_j)}\operatorname{cov}\{\widehat{C}_{\mathbf{s}_1+\mathbf{h}_i,\mathbf{s}_1}(u_i), \widehat{C}_{\mathbf{s}_2+\mathbf{h}_j,\mathbf{s}_2}(u_j)\}.
\end{aligned}$$

□

B.2 Proofs of the theorems

Proof of Theorem 3.2.1

Consider the covariance term, $\operatorname{cov}\{\widehat{C}_n(\mathbf{h}_i, u_i), \widehat{C}_n(\mathbf{h}_j, u_j)\}$, where $(\mathbf{h}_i, u_i)$, $(\mathbf{h}_j, u_j) \in \mathbf{\Lambda}$, and let $D_n^i = D_n(\mathbf{h}_i, u_i)$. Then

$$\begin{aligned}
& \frac{1}{|D_n^i||D_n^j|}\sum_{D_n^i}\sum_{D_n^j}\operatorname{cov}\{Z(\mathbf{s}_1,t_1)Z(\mathbf{s}_1+\mathbf{h}_i,t_1+u_i), Z(\mathbf{s}_2,t_2)Z(\mathbf{s}_2+\mathbf{h}_j,t_1+u_j)\} \\
= \; & \frac{1}{|D_n^i||D_n^j|}\sum_{(\mathbf{s},t)\in D_n^j - D_n^i}\sum_{D_n^i \bigcap (D_n^j-(\mathbf{s},t))}\operatorname{cov}\{Z(\mathbf{0},0)Z(\mathbf{h}_i,u_i), Z(\mathbf{s},t)Z(\mathbf{s}+\mathbf{h}_j,t+u_j)\} \\
= \; & \sum_{D_n^j - D_n^i}\operatorname{cov}\{Z(\mathbf{0},0)Z(\mathbf{h}_i,u_i), Z(\mathbf{s},t)Z(\mathbf{s}+\mathbf{h}_j,t+u_j)\}\frac{|D_n^i \bigcap\{D_n^j-(\mathbf{s},t)\}|}{|D_n^i|\times|D_n^j|}.
\end{aligned}$$

Applying conditions (C2), (3.2) and Kronecker's lemma, we conclude

$$|D_n|\text{cov}\{\widehat{C}_n(\mathbf{h}_i,u_i),\widehat{C}_n(\mathbf{h}_j,u_j)\}\to\sum_{\mathbf{s}\in\mathbb{Z}^2}\sum_{t\in\mathbb{Z}}\text{cov}\{Z(\mathbf{0},0)Z(\mathbf{h}_i,u_i),Z(\mathbf{s},t)Z(\mathbf{s}+\mathbf{h}_j,t+u_j)\}.$$

Let $\sigma^2=\sum_{\mathbf{s}\in\mathbb{Z}^2}\sum_{t\in\mathbb{Z}}\text{cov}\{Z(\mathbf{0},0)Z(\mathbf{h}_1,u_1),Z(\mathbf{s},t)Z(\mathbf{s}+\mathbf{h}_2,t+u_2)\}$, $A_n=\sqrt{|D_n|}\{\widehat{C}_n(\mathbf{h},u)-C_n(\mathbf{h},u)\}$. By Lemma B.1.2, $A_n\xrightarrow{d}\text{N}(0,\sigma^2)$. The proof of the joint normality is analogous to Guan (2003) using the Cramér-Wold device. □

Proof of Corollary 3.2.2

$$\begin{aligned}
&\text{cov}\{\widehat{C}_n(\mathbf{h}_i,u_i),\widehat{C}_n(\mathbf{h}_j,u_j)\}\\
=\;&\tfrac{1}{|S(\mathbf{h}_i)||T_n|}\tfrac{1}{|S(\mathbf{h}_j)||T_n|}\textstyle\sum_{\mathbf{s}_1\in S(\mathbf{h}_i)}\sum_{t_1\in T_n(u_i)}\sum_{\mathbf{s}_2\in S(\mathbf{h}_j)}\sum_{t_2\in T_n(u_j)}\\
&\text{cov}\{Z(\mathbf{s}_1,t_1)Z(\mathbf{s}_1+\mathbf{h}_i,t_1+u_i),Z(\mathbf{s}_2,t_2)Z(\mathbf{s}_2+\mathbf{h}_j,t_2+u_j)\}\\
=\;&\tfrac{1}{|S(\mathbf{h}_i)||S(\mathbf{h}_j)|}\textstyle\sum_{S(\mathbf{h}_i)}\sum_{S(\mathbf{h}_j)}\sum_{T_n(u_j)-T_n(u_i)}\text{cov}\{Z(\mathbf{s}_1,0)Z(\mathbf{s}_1+\mathbf{h}_i,u_i),\\
&Z(\mathbf{s}_2,t)Z(\mathbf{s}_2+\mathbf{h}_j,t+u_j)\}\tfrac{|T_n(u_i)\cap\{T_n(u_j)-t\}|}{|T_n|\times|T_n|}.
\end{aligned}$$

Applying condition (3.5) and Kronecker's Lemma, we conclude

$$\begin{aligned}
&|T_n|\times\text{cov}\{\widehat{C}_n(\mathbf{h}_i,u_i),\widehat{C}_n(\mathbf{h}_j,u_j)\}\\
\to\;&\tfrac{1}{|S(\mathbf{h}_i)|\times S(\mathbf{h}_j)}\sum_{S(\mathbf{h}_i)}\sum_{S(\mathbf{h}_j)}\sum_{t\in\mathbb{Z}}\text{cov}\{Z(\mathbf{s}_1,0)Z(\mathbf{s}_1+\mathbf{h}_i,u_i),Z(\mathbf{s}_2,t)Z(\mathbf{s}_2+\mathbf{h}_j,t+u_j)\}.
\end{aligned}$$

Letting $A_n=\sqrt{|T_n|}\times\{\widehat{C}_n(\mathbf{h},u)-C_n(\mathbf{h},u)\}$, we prove $A_n\xrightarrow{\quad}\text{N}(0,\sigma^2)$ by applying Lemma B.1.4. Specifically, in this theorem $p=2,q=1$ and the mixing condition (3.3) does not depend on the cardinality, which is a special case of (C2) in Lemma B.1.4. Consequently, here we choose α and η using condition $1/(1+\epsilon)<\eta<\alpha<1$, and

$$|\phi_n'(x)-\phi_n(x)|\le\sum_{j=1}^{k_n-1}16\alpha(n^\eta)=\text{O}(n^{1-\alpha}n^{-\epsilon\eta})=\text{O}(n^{1-\alpha-\epsilon\eta}).$$

$1-\alpha-\epsilon\eta<0$ since $\frac{1}{1+\epsilon}<\eta<\alpha<1$. Thus $|\phi_n'(x)-\phi_n(x)|\to 0$ as $n\to\infty$.

The Cramér-Wold's device proves the joint normality. □

Proof of Theorem 3.3.1

It is a special case of Lemma B.1.5 setting $p = 2$ and $q = 1$. □

VITA

Bo Li received a bachelor of science degree in Naval Architecture & Ocean Engineering and a master of science degree in Ship & Ocean Engineering Fluid Mechanics from Shanghai Jiaotong University. After that, she worked as a Naval Architect for two years and then she was admitted to the master program in the Department of Statistics at Texas A&M University. She received her master of science degree in statistics in May 2004, and continued to study in the Ph.D. program in this department. She received her Ph.D. degree in 2006. Her permanent address is: 17 Xiangyang Road, Apt. 13-203, Anqiu, Shandong, People's Republic of China.